Probability and Statistics - P_{31}

药品试验问题，假设对于治疗某种疾病而研制的两种新药

设新药 $i=1,2$ 的治愈率为 P_i，即接受药品 i 治疗的病人的治愈概率为 P_i

假定对于不同病人 P_i 的差异可以忽略，且相互之间是独立的

而 P_i 是未知的，则考虑判断 $P_1>P_2$ 或 $P_1<P_2$

试验是成对地，有序地进行的，即对于每一对病人，分别用药品1和2治疗

当其中一种药品治愈人数超过另一种药品的治愈人数一定数量 M 时即停止

令指示器变量 $X_j = \begin{cases} 1, & \text{第}j\text{对试验中药品1治愈了病人} \\ 0, & \text{其他} \end{cases}$

$Y_j = \begin{cases} 1, & \text{第}j\text{对试验中药品2治愈了病人} \\ 0, & \text{其他} \end{cases}$

令随机变量 N 为对于指定的常数 M

等式 $(X_1+X_2+\cdots+X_n)-(Y_1+Y_2+\cdots+Y_n)=M$

或 $(X_1+X_2+\cdots+X_n)-(Y_1+Y_2+\cdots+Y_n)=-M$ 首次成立时 n 的值

考虑试验的正确性，即当给定 $P_1>P_2$ 的前提下，做出 $P_1<P_2$ 的结论的可能

注意到药品1和2同时有效或同时无效时，累计差保持不变

概率为 $P_1P_2+(1-P_1)(1-P_2)$

当药品1有效而药品2无效时，累计差 $+1$

概率为 $P_1(1-P_2)$

当药品1无效而药品2有效时，累计差 -1

概率为 $(1-P_1)P_2$

如果只考虑使累计差产生变化的试验

令概率 $P = Pr\{\text{累计差}+1 \mid \text{累计差变化}\} = \dfrac{P_1(1-P_2)}{P_1(1-P_2)+(1-P_1)P_2}$

则 $1-P = Pr\{\text{累计差}-1 \mid \text{累计差变化}\} = \dfrac{(1-P_1)P_2}{P_1(1-P_2)+(1-P_1)P_2}$

于是做出 $P_1<P_2$ 的结论可以描述为，以概率 P 赢得1元的赌徒破产问题

在累计 $+M$ 之前出现累计 $-M$ 的概率

则有 $Pr\{\text{结论为}P_2>P_1\} = 1-[1-(\frac{1-P}{P})^M]/[1-(\frac{1-P}{P})^{2M}]$

$= 1-[1-(\frac{1-P}{P})^M]/[1-(\frac{1-P}{P})^M][1+(\frac{1-P}{P})^M]$

$= 1-1/[1+(\frac{1-P}{P})^M]$

$= (\frac{1-P}{P})^M/[1+(\frac{1-P}{P})^M]$

$= 1/[1+(\frac{P}{1-P})^M]$

取 $\lambda = \dfrac{P}{1-P} = \dfrac{P_1(1-P_2)}{(1-P_1)P_2}$，则 $Pr\{\text{结论为}P_1<P_2\} = \dfrac{1}{1+\lambda^M}$

于是当 $P_1>P_2$ 时，$\lambda>1$，则有 $\lim\limits_{M\to+\infty} Pr\{\text{结论为}P_1<P_2\}=0$

如对于 $P_1=0.6$，$P_2=0.4$，$\lambda=2.25$，

当 $M=5$ 时，$Pr\{P_1<P_2\}=0.017$，而当 $M=10$ 时，$Pr\{P_1<P_2\}=0.0003$

Probability and Statistics - P32

条件概率公理：对于样本空间 S 上的事件 E 和 F，以及互不相容的事件序列 $E_i\ (i=1,2,\cdots)$

有 $0 \le P(E|F) \le 1$

$P(S|F) = 1$

$P(\bigcup_{i=1}^{\infty} E_i | F) = \sum_{i=1}^{\infty} P(E_i|F)$

证明过程有，对于 $0 \le P(E|F) \le 1$

当 $P(F)=0$ 时，有 $P(E|F)=0$

当 $P(F)>0$ 时，有 $P(E|F) = P(EF)/P(F)$

又由于 $0 \le P(EF)$，于是 $P(E|F) = P(EF)/P(F) \ge 0$

而 $EF \subseteq F$，于是有 $P(EF) \le P(F)$

则有 $0 \le P(E|F) \le 1$

以下假定 $P(F)>0$，对于 $P(S|F)=1$

由于 $F \subseteq S$，则有 $SF = F$

于是 $P(S|F) = P(SF)/P(F) = P(F)/P(F) = 1$

对于 $P(\bigcup_{i=1}^{\infty} E_i | F) = \sum_{i=1}^{\infty} P(E_i|F)$

有 $P(\bigcup_{i=1}^{\infty} E_i|F) = P[(\bigcup_{i=1}^{\infty} E_i)F]/P(F)$

根据分配率 $= P(\bigcup_{i=1}^{\infty} E_iF)/P(F)$

由于 $E_i\ (i=1,2,\cdots)$ 是互不相容事件 $= [\sum_{i=1}^{\infty} P(E_iF)]/P(F)$

于是 $E_iF\ (i=1,2,\cdots)$ 也是互不相容事件 $= \sum_{i=1}^{\infty} \frac{P(E_iF)}{P(F)}$

$= \sum_{i=1}^{\infty} P(E_i|F)$

条件容斥原理：对于样本空间 S 上的事件 E_1, E_2, F，其中 $P(F)>0$

于是有 $P(E_1 \cup E_2 | F) = P(E_1|F) + P(E_2|F) - P(E_1E_2|F)$

条件贝叶斯公式：对于样本空间 S 上的事件 H, E, F，其中 $P(F)>0$，$P(EF)>0$，$P(E^cF)>0$

于是有 $P(H|F) = P(H|EF)P(E|F) + P(H|E^cF)P(E^c|F)$

证明过程有，
$$P(H|F) = \frac{P(HF)}{P(F)} = \frac{P(HEF)+P(HE^cF)}{P(F)}$$
$$= \frac{P(HEF)}{P(F)} + \frac{P(HE^cF)}{P(F)}$$
$$= \frac{P(HEF)}{P(EF)} \cdot \frac{P(EF)}{P(F)} + \frac{P(HE^cF)}{P(E^cF)} \cdot \frac{P(E^cF)}{P(F)}$$
$$= P(H|EF)P(E|F) + P(H|E^cF)P(E^c|F)$$

条件独立 (conditionally independent)：对于样本空间 S 上的事件 E_1, E_2, F

如果有 $P(E_1|E_2F) = P(E_1|F)$ 或等价地有 $P(E_1E_2|F) = P(E_1|F)P(E_2|F)$

则称事件 E_1, E_2 是关于事件 F 是条件独立的

Probability and Statistics - P33

游程

对于一个独立重复试验的序列，成功的概率为 p，失败的概率为 $q=1-p$

则考虑长度为[redacted] n 的成功游程先于长度为 m 的失败游程出现的概率

令事件 E 为长度为 n 的成功游程先于长度为 m 的失败游程出现

事件 H 表示第一次试验成功，

于是有 $Pr(E) = Pr(E|H)Pr(H) + Pr(E|H^c)Pr(H^c)$

$= p\,Pr(E|H) + q\,Pr(E|H^c)$

考虑当第一次试验成功，即事件 H 发生

为了使长度为 n 的成功游程先于长度为 m 的失败游程出现

则需要接下去 $n-1$ 次试验均成功

令事件 F 表示第2次到第 n 次试验均成功

则根据条件[redacted]贝叶斯公式

则有 $Pr(E|H) = Pr(E|FH)Pr(F|H) + Pr(E|F^cH)Pr(F^c|H)$

对于 $Pr(E|FH)$，显然有 $Pr(E|FH)=1$

而对于 $Pr(E|F^cH)$，表示第2次到第 n 次试验中至少有1次失败

由于独立重试验的无记忆性，则失败发生时之前的成功即失效

[redacted]于是有 $Pr(E|F^cH) = Pr(E|H^c)$

又 F 与 H 是独立的，即有 $Pr(F|H) = Pr(F)$，$Pr(F^c|H) = Pr(F^c)$

于是 $Pr(E|H) = Pr(F) + Pr(F^c)Pr(E|H^c)$

$= p^{n-1} + (1-p^{n-1})Pr(E|H^c)$

再考虑第一次试验失败，即事件 H^c 发生

令事件 G 表示第2次到第 m 次试验均失败

则对于 $Pr(E|GH^c)$，显然有 $Pr(E|GH^c)=0$

而对于 $Pr(E|G^cH^c)$，与 $Pr(E|F^cH)$ 类似地有 $Pr(E|G^cH^c)=Pr(E|H)$

于是有 $Pr(E|H^c) = Pr(E|GH^c)P(G|H^c) + Pr(E|G^cH^c)P(G^c|H^c)$

$= Pr(E|H)P(G^c) = (1-q^{m-1})Pr(E|H)$

于是有 $\begin{cases} Pr(E|H) = p^{n-1} + (1-p^{n-1})Pr(E|H^c) \\ Pr(E|H^c) = (1-q^{m-1})Pr(E|H) \end{cases}$

解之得 $\begin{cases} Pr(E|H) = p^{n-1}/(p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}) \\ Pr(E|H^c) = p^{n-1}(1-q^{m-1})/(p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}) \end{cases}$

则 $Pr(E) = p\,Pr(E|H)+q\,Pr(E|H^c) = \frac{p^{n-1}(p+q-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}} = \frac{p^{n-1}(1-q^m)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$

由于问题的对称性，交换 n 与 m，p 与 q

则有 $Pr($长度为 m 的失败序列先于长度为 n 的成功序列出现$) = \frac{q^{m-1}(1-p^n)}{p^{n-1}+q^{m-1}-p^{n-1}q^{m-1}}$

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

即长度为 n 的成功序列与长度为 m 的失败序列终有一个会出现

Probability and Statistics - P34

对于相互独立的事件序列 $E_1, E_2, \cdots, E_n$

有 $P(E_1 \cup E_2 \cup \cdots \cup E_n) = 1 - \prod_{i=1}^{n} [1 - P(E_i)]$

证明过程有，考虑事件 $E_1 \cup E_2 \cup \cdots \cup E_n$ 的补

则有 $(E_1 \cup E_2 \cup \cdots \cup E_n)^c = E_1^c \cap E_2^c \cap \cdots \cap E_n^c$

又 $E_1, E_2, \cdots, E_n$ 是相互独立的

则 $E_1^c, E_2^c, \cdots, E_n^c$ 也是相互独立的

于是
$$\begin{aligned} P(E_1 \cup E_2 \cup \cdots \cup E_n) &= 1 - P[(E_1 \cup E_2 \cup \cdots \cup E_n)^c] \\ &= 1 - P(E_1^c \cap E_2^c \cap \cdots \cap E_n^c) \\ &= 1 - P(E_1^c) P(E_2^c) \cdots P(E_n^c) \\ &= 1 - \prod_{i=1}^{n} P(E_i^c) \\ &= 1 - \prod_{i=1}^{n} [1 - P(E_i)] \end{aligned}$$

特别地有，假设有无限多个相互独立的事件序列 $E_1, E_2, \cdots$

则有 $P(\bigcup_{i=1}^{\infty} E_i) = 1 - \prod_{i=1}^{\infty} [1 - P(E_i)]$

对于无限多个实数 $0 \le a_i \le 1$，$i = 1, 2, \cdots$

有 $\sum_{i=1}^{\infty} [a_i \prod_{j=1}^{i-1} (1 - a_j)] + \prod_{i=1}^{\infty} (1 - a_i) = 1$

证明过程有，考虑一个掷无限多个硬币的过程，硬币之间是相互独立

令实数 $0 \le a_i \le 1$ 表示第 i 枚硬币正面向上的概率，$i = 1, 2, \cdots$

令事件 E_i 表示第 i 枚硬币是第一枚正面向上的硬币，$i = 1, 2, \cdots$

即前 $i-1$ 枚硬币均为反面

令事件 C_i 表示第 i 枚硬币正面向上，则 C_i 为相互独立的事件序列

且有 $P(C_i) = a_i$，$P(C_i^c) = 1 - a_i$

注意到有 $E_i = C_1^c C_2^c \cdots C_{i-1}^c C_i$，$i = 1, 2, \cdots$

于是 E_i 也是相互独立的事件序列

且有事件 $\bigcap_{i=1}^{\infty} E_i^c$ 表示所有硬币均不是第一枚正面向上的硬币

于是有 $\bigcap_{i=1}^{\infty} E_i^c = \bigcap_{i=1}^{\infty} C_i^c$

又 E_i 同时是互不相交的事件序列

于是 $P(\bigcup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} P(E_i)$

$P(\bigcap_{i=1}^{\infty} E_i^c) = P(\bigcap_{i=1}^{\infty} C_i^c)$

又 $P(\bigcup_{i=1}^{\infty} E_i) = 1 - P(\bigcap_{i=1}^{\infty} E_i^c)$

$\sum_{i=1}^{\infty} P(E_i) = 1 - P(\bigcap_{i=1}^{\infty} C_i^c)$

$\sum_{i=1}^{\infty} P(C_i \bigcap_{j=1}^{i-1} C_j^c) = 1 - \prod_{i=1}^{\infty} P(C_i^c)$

$\sum_{i=1}^{\infty} [a_i \prod_{j=1}^{i-1} (1 - a_j)] = 1 - \prod_{i=1}^{\infty} (1 - a_i)$，即 $\sum_{i=1}^{\infty} [a_i \prod_{j=1}^{i-1} (1 - a_j)] + \prod_{i=1}^{\infty} (1 - a_i) = 1$

Probability and Statistics - P35

伯努利试验序列 (Bernoulli trial)：对于独立重复试验序列中，每次试验结果只有成功/失败

对于一个 n 次独立重复的伯努利试验序列，$n \in N$

每次试验成功的概率为 p，则失败的概率为 $1-p$

考虑试验序列中出现偶数次成功的概率，认为0次也是偶数次

令 P_n 表示 n 次伯努利试验序列中出现偶数次成功的概率

令事件 E_n 表示 n 次伯努利试验序列中出现偶数次成功

则当 $n=0$ 时，平凡地有 $P_n = Pr(E_n) = 1$

当 $n>0$ 时，考虑最后一次试验的结果

令事件 H 表示最后一次试验结果为成功

于是 $Pr(E_n) = Pr(E_n | H)P(H) + Pr(E_n | H^c)P(H)$

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

$+ [Pr(E_n | E_{n-1}H^c)P(E_{n-1}|H^c) + Pr(E_n | E_{n-1}^c H^c)P(E_{n-1}^c | H^c)]P(H^c)$

又 $Pr(E_n | E_{n-1}H) = Pr(E_n | E_{n-1}^c H^c) = 0$，$Pr(E_n | E_{n-1}H^c) = Pr(E_n | E_{n-1}^c H) = 1$

E_{n-1} 与 H 相互独立，$P(H) = p$，$P(H^c) = (1-p)$

于是 $Pr(E_n) = P(E_{n-1}^c)P(H) + P(E_{n-1})P(H^c)$

即 $P_n = p(1-P_{n-1}) + (1-p)P_{n-1}$，$n \geq 1$

于是 n 次伯努利试验序列中出现偶数次成功的概率服从递推关系

$P_0 = 1$，$P_n = p(1-P_{n-1}) + (1-p)P_{n-1}$，$n \geq 1$，其中 p 为每次试验的成功概率

对于一个每次试验成功率为 p 的 n 次伯努利试验序列，$n \in N$，$0 \leq p \leq 1$

序列中出现偶数次成功的概率 $P_n = \frac{1+(1-2p)^n}{2}$

基础步骤：当 $n=0$ 时，由初始条件可知 $P_0 = 1$

$\frac{1+(1-2p)^n}{2} = \frac{1+1}{2} = 1 = P_0$

递归步骤：假设对于任意 $n>0$，$P_{n-1} = \frac{1+(1-2p)^{n-1}}{2}$ 成立

则考虑 P_n，由递推关系可知

$P_n = p(1-P_{n-1}) + (1-p)P_{n-1}$

$= p + (1-2p)P_{n-1}$

$\overset{(IH)}{=} p + (1-2p)\frac{1+(1-2p)^{n-1}}{2}$

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

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

于是根据数学归纳法：对于每次试验成功率为 p 的 n 次伯努利试验序列

序列中出现偶数次成功的概率 $P_n = \frac{1+(1-2p)^n}{2}$

特别地有，当 $p=\frac{1}{2}$ 时，对于任意 $n>0$，有 $P_n = \frac{1}{2}$

Probability and Statistics - P36

无限赌本

假设与一个无限富裕的人赌[redacted]博，每一次[redacted]或者赢得1单位，或者输掉1单位

如果有n单位初始资金，其中 $n \in N$，

每一次赌博都是独立且等概率获胜的，赢的概率为p，输的概率为 $q = 1-p$

则考虑对于给定的初始资金n和胜率p，最终输光的概率

令 P_n 表示初始资金为n单位时的输光的概率

当 $p=0$ 时，对于任意 $n \in N$，由于胜率为0

则有 P_n 平凡地等于1

则考虑胜率 $0 < p \le 1$ 时的 P_n

当 $n=0$ 时，由于此时已处于输光的状态，于是 $P_0 = 1$

当 $n \ge 1$ 时，考虑第一次赌博的结果

令事件H表示第一次赌博获胜，E_n 表示初始为n单位且最终输光

于是有 $P_n = Pr(E_n) = Pr(E_n | H) P(H) + Pr(E_n | H^c) P(H^c)$

$= Pr(E_{n+1}) P(H) + P(E_{n-1}) P(H_c)$

$= p \cdot P_{n+1} + q \cdot P_{n-1}$

于是可知 P_n 满足递推关系 $P_n = p \cdot P_{n+1} + q \cdot P_{n-1}$

取特征方程 $r = p \cdot r^2 + q$，即 $pr^2 - r + (1-p) = 0$

于是有 $r_1 = \frac{1-p}{p} = q/p$，$r_2 = 1$

则递推关系 $P_n = p \cdot P_{n+1} + q \cdot P_{n-1}$ 的解为

$P_n = \alpha_1 (q/p)^n + \alpha_2$，其中 α_1, α_2 为常数

注意到当[redacted] $p < \frac{1}{2}$ 时，$q/p > 1$，又对于任意n应有 $0 \le P_n \le 1$

于是取 $\alpha_1 = 0$，$\alpha_2 = 1$，则有 $P_n = 1$

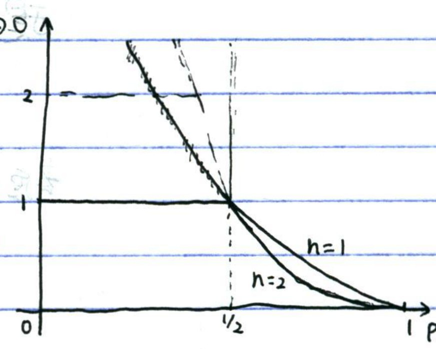
而当 $p \ge \frac{1}{2}$ 时，$q/p \le 1$，于是常数 α_1 可不为0

[redacted]

于是取 $\alpha_1 = 1$，$\alpha_2 = 0$，则有 $P_n = (q/p)^n$

于是可知对于初始资金为n单位

有 $P_n = \begin{cases} 1, & p \le \frac{1}{2} \\ (\frac{q}{p})^n, & p > \frac{1}{2} \end{cases}$



另一种解释为，令事件E为[redacted]任意一次赌博之后，总结果为输掉1单位赌本

则可知 $P_n = P(\cap_{i=1}^{n} E) = \prod_{i=1}^{n} P(E) = [P(E)]^n$，即积地输掉n单位赌本

则 $P(E) = q + p[P(E)]^2$，即 $P(E) = q/p$ 或 $P(E) = 1$

于是当 $p \le \frac{1}{2}$ 时，$P(E) = 1$，即 $P_n = 1$

当 $p > \frac{1}{2}$ 时，$P(E) = q/p$ 即 $P_n = (\frac{q}{p})^n$

Probability and Statistics - P37

拉普拉斯继承准则 (Laplace's rule of succession)，源于Laplace解日出问题 (sunrise problem)

对于一个有 $k+1$ 枚硬币的盒子，从中随机取一枚硬币，并重复抛掷

对于第 i 枚硬币，每次抛掷的独立且正面概率均为 i/k，$i=0,1,\cdots,k$，$k\in N^+$

如果前 n 次抛掷均为正面，则考虑第 $n+1$ 次抛掷仍是正面向上的概率

令事件 C_i ($i=0,1,\cdots,k$) 表示取出的是第 i 枚硬币

事件 F_n 表示前 n 次抛掷结果均为正面，$n\in N$

事件 H 表示第 $n+1$ 次抛掷结果为正面

于是有 $P(H|F_n)=\sum_{i=0}^{k}P(H|F_nC_i)P(C_i|F_n)$

已知取出的是第 i 枚硬币，于是可以认为每次抛掷是条件独立的

注意 事件 E_1,E_2 关于给定事件 F 独立

即 $E_1\perp E_2|F \longleftrightarrow P(E_1E_2|F)=P(E_1|F)P(E_2|F)$

于是 $P(E_1|F)=\dfrac{P(E_1E_2|F)}{P(E_2|F)}=\dfrac{P(E_1E_2F)/P(F)}{P(E_2F)/P(F)}$

$=P(E_1E_2F)/P(E_2F)$

$=P(E_1|E_2F)$

于是 $E_1\perp E_2|F \longleftrightarrow P(E_1|E_2F)=P(E_1|F)$

即有 $P(H|F_nC_i)=P(H|C_i)=i/k$

又 $P(C_i|F_n)=\dfrac{P(C_iF_n)}{P(F_n)}=\dfrac{P(F_n|C_i)P(C_i)}{\sum_{j=0}^{k}P(F_n|C_j)P(C_j)}$

又随机地取出一枚硬币，即 $P(C_i)=\dfrac{1}{1+k}$，$i=0,1,\cdots,k$

$P(F_n|C_i)=(i/k)^n$，$i=0,1,\cdots,k$

于是 $P(C_i|F_n)=\dfrac{P(F_n|C_i)\frac{1}{k+1}}{\sum_{j=0}^{k}P(F_n|C_j)\frac{1}{1+k}}=\dfrac{P(F_n|C_i)}{\sum_{j=0}^{k}P(F_n|C_j)}$

$=(i/k)^n/[\sum_{j=0}^{k}(j/k)^n]$

于是 $P(H|F_n)=\sum_{i=0}^{k}P(H|C_i)P(C_i|F_n)$

$=\sum_{i=0}^{k}(i/k)\cdot(i/k)^n/[\sum_{j=0}^{k}(j/k)^n]$

$=[\sum_{i=0}^{k}(i/k)^{n+1}]/[\sum_{j=0}^{k}(j/k)^n]$

考虑当 k 足够大的情形，此时求和可以看作积分

即 $\lim_{k\to+\infty}\frac{1}{k}\sum_{i=0}^{k}(i/k)^{n+1}=\int_0^1 x^{n+1}dx=\dfrac{1}{n+2}$

$\lim_{k\to+\infty}\frac{1}{k}\sum_{i=0}^{k}(i/k)^{n}=\int_0^1 x^{n}dx=\dfrac{1}{n+1}$

此处取 $\Delta x=\frac{1}{k}$

于是可知对于足够大的 $k\in N^+$

有 $\lim_{k\to+\infty}P(H|F_n)=\dfrac{n+1}{n+2}$

在过去的45亿年中太阳每天都升起，则有 $n=45$亿年 $=1.6425\times10^{12}$ 天

则 $P(\text{太阳明天升起})=\dfrac{n+1}{n+2}\approx 1-10^{-13}$

Probability and Statistics - P38

条件全概率公式，对于 n 个互不相容且穷举的假设 $H_1, H_2, \cdots, H_n$

即有对于任意 $1 \le i < j \le n$，有 $H_i \cap H_j = \emptyset$

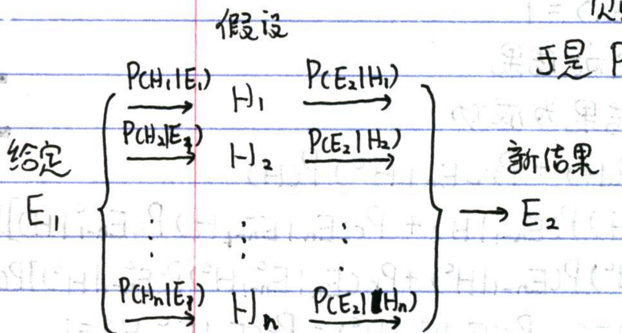
且 $H_1 \cup H_2 \cup \cdots \cup H_n = S$，其中 S 为样本空间

如果有事件 E_1, E_2 对于假设 H_i 条件独立，其中 $i = 1, 2, \cdots, n$

则有 $P(E_2|E_1) = \sum_{i=1}^n P(E_2|H_i)P(H_i|E_1)$

证明过程有，由于事件 E_1, E_2 关于假设 H_i 条件独立，$i = 1, 2, \cdots, n$

则有 $P(E_1E_2|H_i) = P(E_1|H_i)P(E_2|H_i)$



于是 $P(E_2|E_1) = \frac{P(E_1E_2)}{P(E_1)} = P(\cup_{i=1}^n H_iE_1E_2)/P(E_1)$

$= \sum_{i=1}^n [P(H_iE_1E_2)/P(E_1)]$

$= \sum_{i=1}^n [P(E_1E_2|H_i)P(H_i)/P(E_1)]$

$= \sum_{i=1}^n [P(E_1|H_i)P(E_2|H_i)P(H_i)/P(E_1)]$

$= \sum_{i=1}^n [P(E_2|H_i)P(E_1H_i)/P(E_1)]$

$= \sum_{i=1}^n P(E_2|H_i)P(H_i|E_1)$

也可以描述为在给定事件 E_1 发生的条件下 E_2 发生的概率

可以分解为给定事件 E_1 发生，则 n 个互不相容且穷举的假设之一 H_i 成立

且在该假设 H_i 成立的条件下，事件 E_2 发生

对于 n 个互不相容且穷举的假设 $H_1, H_2, \cdots, H_n$ 与先后发生的事件 E_1, E_2

先验概率 (prior probability)，也称初始概率，为事件发生前的概率分布，即 $\sum_{i=1}^n P(H_i) = 1$

后验概率 (posterior probability)，也称条件概率，为事件发生后考虑相关证据的条件概率分布

即 $P(H_i|E_1) = \frac{P(E_1|H_i)P(H_i)}{P(E_1)} = \frac{P(E_1|H_i)P(H_i)}{\sum_{j=1}^n P(E_1|H_j)P(H_j)}$，即有 $\sum_{i=1}^n P(H_i|E_1) = 1$

则当 E_2 发生后，此时拥有两个事件发生的信息

于是此时的后验概率为 $P(H_i|E_1E_2) = P(E_1E_2|H_i)P(H_i)/[\sum_{j=1} P(E_1E_2|H_j)P(H_j)]$

序贯地补充信息　考虑当 E_1, E_2 关于任意假设 H_i，$i = 1, 2, \cdots, n$ 都是条件独立的

则 $P(H_i|E_1E_2) = \frac{P(E_1E_2|H_i)P(H_i)}{P(E_1E_2)} = \frac{P(E_1|H_i)P(E_2|H_i)P(H_i)}{P(E_1E_2)}$

$= P(E_2|H_i)P(E_1H_i)/P(E_1E_2) = \frac{P(E_2|H_i)P(H_i|E_1)/P(E_1)}{P(E_1E_2)/P(E_1)}$

$= \frac{P(E_2|H_i)P(H_i|E_1)}{P(E_2|E_1)} = P(E_2|H_i)P(H_i|E_1)/[\sum_{j=1}^n P(E_2|H_j)P(H_j|E_1)]$

于是可以描述为，当条件独立的事件 E_1, E_2 先后发生时

在 E_1 发生后 E_2 发生前可视 $P(H_i|E_1)$ 为"最新的"先验概率

通过对 $P(H_i|E_1)$ 进行修正以得到"最新的"后验概率 $P(H_i|E_1E_2)$

扩展至一系列的关于假设条件独立的事件 $E_1, E_2, \cdots, E_k, \cdots$

都可以将 $P(H_i|E_1E_2\cdots E_{k-1})$ 视为先验概率，而仅经过 $P(E_k|H_i)$ 调整得到后验概率

体现了一种无记忆性

Probability and Statistics - P39

拉普拉斯继承准则扩展，对于一个有k+1枚硬币的盒子，从中随机取一枚，并重复地抛掷

对于第 i 枚硬币，$i=0,1,\cdots,k$，每次抛掷独立且正面概率为 i/k

如果前n次抛掷中有 $0\le s\le n$ 次正面，则求第n+1次为正面的概率

令事件 C_i 表示取出的第i枚硬币，其中 $i=0,1,\cdots,k$

事件 $F_n=\{x_1,x_2,\cdots,x_n\}$，其中 x_j 表示第j次抛掷的结果

令 $x_j=1$ 表示正面，$x_j=0$ 表示反面，则 $\sum_{j=1}^n x_i = s$ 表示有 $0\le s\le n$ 次正面

事件H表示第n+1次抛掷的结果为正面

于是有 $P(H|F_n)=\sum_{i=0}^k P(H|F_nC_i)P(C_i|F_n)$

由于抛掷之间是条件独立的，于是有 $P(H|F_nC_i)=P(H|C_i)=i/k$

而 $P(C_i|F_n)=\dfrac{P(C_iF_n)}{P(F_n)}=\dfrac{P(F_n|C_i)P(C_i)}{\sum_{l=0}^k P(F_n|C_l)P(C_l)}$

又硬币为随机抽取的，于是有 $P(C_l)=\frac{1}{k+1}$，$l=0,1,\cdots,k$

$$P(C_i|F_n)=\frac{P(F_n|C_i)\cdot\frac{1}{1+k}}{\sum_{l=0}^k P(F_n|C_l)\cdot\frac{1}{1+k}}=\frac{P(F_n|C_i)}{\sum_{l=0}^k P(F_n|C_l)}$$

而 $P(F_n|C_i)=\prod_{j=1}^n (i/k)^{x_j}(1-i/k)^{1-x_j}$，而 $\sum_{j=1}^n x_j=s$

$$=(i/k)^s(1-i/k)^{n-s}$$

于是 $P(C_i|F_n)=\dfrac{(i/k)^s(1-i/k)^{n-s}}{\sum_{l=0}^k (l/k)^s(1-l/k)^{n-s}}$，其中 $0\le s\le n$

即 $P(H|F_n)=\sum_{i=0}^k\left[(i/k)^{s+1}(1-i/k)^{n-s}/\sum_{l=0}^k (l/k)^s(1-l/k)^{n-s}\right]$

$$=\sum_{i=0}^k (i/k)^{s+1}(1-i/k)^{n-s}\Big/\sum_{l=0}^k (l/k)^s(1-l/k)^{n-s}$$

当k足够大时，求和式可以看作积分

即 $\lim\limits_{k\to\infty}\frac{1}{k}\sum_{i=0}^k (i/k)^{s+1}(1-i/k)^{n-s}=\int_0^1\theta^{s+1}(1-\theta)^{n-s}d\theta=\dfrac{\Gamma(s+2)\Gamma(n-s+1)}{\Gamma(n+3)}$

$\lim\limits_{k\to\infty}\frac{1}{k}\sum_{i=0}^k (i/k)^{s}(1-i/k)^{n-s}=\int_0^1\theta^{s}(1-\theta)^{n-s}d\theta=\dfrac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)}$

于是有 $\lim\limits_{k\to\infty}P(H|F_n)=\lim\limits_{k\to\infty}\frac{1}{k}\sum_{i=0}^k (i/k)^{s+1}(1-i/k)^{n-s}\Big/\frac{1}{k}\sum_{i=0}^k (i/k)^s(1-i/k)^{n-s}$

$$=\frac{\Gamma(s+2)\Gamma(n-s+1)}{\Gamma(n+3)}\Big/\frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)}$$

$$=\frac{\Gamma(s+2)/\Gamma(s+1)}{\Gamma(n+3)/\Gamma(n+2)}=\frac{s+1}{n+2}$$

拉普拉斯继承准则可以进一步扩展为连续概率分布的情形

即随机地选择一个概率 $0\le p\le 1$

则有 $P(H|F_n)=\int_0^1\dfrac{P(F_n|p)p}{\int_0^1 P(F_n|p_0)dp_0}dp$

$$=\int_0^1 p^{s+1}(1-p)^{n-s}dp\Big/\int_0^1 p_0^s(1-p_0)^{n-s}dp_0$$

$$=\frac{\Gamma(s+2)\Gamma(n-s+1)}{\Gamma(n+3)}\Big/\frac{\Gamma(s+1)\Gamma(n-s+1)}{\Gamma(n+2)}=\frac{s+1}{n+2}$$

如果选择的概率 $0\le\theta\le 1$ 服从概率密度函数 $f(\theta)$

则有 $P(H|F_n)=\int_0^1\dfrac{P(F_n|\theta)\cdot\theta\cdot f(\theta)}{\int_0^1 P(F_n|\theta_0)f(\theta_0)d\theta_0}d\theta$

$$=\int_0^1 P(F_n|\theta)\cdot\theta\cdot f(\theta)d\theta\Big/\int_0^1 P(F_n|\theta_0)f(\theta_0)d\theta_0$$

$$=\int_0^1\theta^{s+1}(1-\theta)^{n-s}f(\theta)d\theta\Big/\int_0^1\theta_0^s(1-\theta_0)^{n-s}f(\theta_0)d\theta_0$$

Probability and Statistics - P40

对于非负整数 m, n，有 $\int_0^1 y^n(1-y)^m dy = \frac{n!m!}{(n+m+1)!}$，其中 $0<y<1$

证明过程有，令 $S_{n,m} = \int_0^1 y^n(1-y)^m dy$，

基础步骤：当 $n=0, m=0$ 时

$$S_{0,0} = \int_0^1 1\,dy = 1 = \frac{0!\cdot 0!}{1!}$$ ，即 $P(0,0)$ 为真

当 $n=0, m>0$ 时

$$S_{0,m} = \int_0^1 (1-y)^m dy = -\frac{1}{m+1}(1-y)^{m+1}\Big|_0^1$$
$$= \frac{1}{m+1} = \frac{0!\cdot m!}{(m+1)!}$$ ，即 $P(0,m)$ 为真

当 $n>0, m=0$ 时

$$S_{n,0} = \int_0^1 y^n dy = \frac{1}{n+1}y^{n+1}\Big|_0^1$$
$$= \frac{1}{n+1} = \frac{n!\cdot 0!}{(n+1)!}$$ ，即 $P(n,0)$ 为真

递归步骤：假设对于任意 $n>0, m>0$

$$S_{n,m} = \int_0^1 y^n(1-y)^m dy$$
$$= \int_0^1 \frac{1}{n+1}(1-y)^m d(y^{n+1})$$
$$= \frac{1}{n+1}(1-y)^m y^{n+1}\Big|_0^1 - \int_0^1 y^{n+1}\, d\left[\frac{1}{n+1}(1-y)^m\right]$$
$$= 0 + \int_0^1 y^{n+1}\cdot\frac{m}{n+1}(1-y)^{m-1}dy$$
$$= \frac{m}{n+1}S_{n+1,m-1}$$
$$= \frac{m}{n+1}\times\frac{m-1}{n+2}\times\cdots\times\frac{1}{n+m}\times S_{n+m,0}$$
$$\overset{(IH)}{=} \frac{m!}{(n+1)(n+2)\cdots(n+m)}\cdot\frac{1}{n+m+1}$$
$$= \frac{n!m!}{(n+m+1)!}$$ ，即 $P(n,m)$ 为真

于是可知 对于任意非负整数 n, m

有 $\int_0^1 y^n(1-y)^m dy = S_{n,m} = \frac{n!m!}{(n+m+1)!}$

注意这个结论也可以用于扩展的拉普拉斯继承准则的推导

拉普拉斯继承准则，对于一枚正面向上概率为 $0<p<1$ 的硬币

其中概率 p 服从 $(0,1)$ 上的均匀分布，且每次投掷间相互独立

则连续地抛掷硬币 n 次且全部正面向上，

考虑在接下来的 m 次抛掷中也是全部正面的条件概率

有 $$P(H_m|F_n) = \int_0^1 \frac{P(F_n|p)p^m}{\int_0^1 P(F_n|p)dp}dp$$
$$= \int_0^1 p^{n+m}(1-p)^0 dp \Big/ \int_0^1 p^n(1-p)^0 dp$$
$$= \frac{(n+m)!\cdot 0!}{(n+m+1)!} \Big/ \frac{n!\cdot 0!}{(n+1)!}$$
$$= \frac{n+1}{n+m+1}$$

则前 n 次抛掷均为正面已知，后 m 次抛掷也均为正面的条件概率为 $\frac{n+1}{n+m+1}$

Probability and Statistics - P41

投票问题

假设在一次选举中，A获得了n张选票，B获得了m张选票，且有 $0\le m<n$

假定在计票过程中 $\binom{n+m}{n}$ 种唱票顺序是等可能出现的

考虑在唱票过程中A始终不落后于B的概率

令 $P_{n,m}$ 表示在统计选票中A始终不落后的概率，其中 $0\le m\le n$

首先考虑边界条件，即当 $m=n$ 时的情形

注意到在 $\binom{2n}{n}$ 种唱票顺序中，A始终不落后的顺序符合第n个卡塔兰数的定义

即有 $\binom{2n}{n}-\binom{2n}{n+1}=\frac{1}{n+1}\binom{2n}{n}$ 种

于是当 $n=m\ge 0$ 时，有 $P_{n,m}=\frac{1}{n+1}$

再考虑当 $0\le m<n$ 时的 $P_{n,m}$

考虑 $P_{n,m}$ 与 $P_{n-1,m}$ 和 $P_{n,m-1}$ 的关系

注意由于 $n>m$，于是 $n-1\ge m$，$n>m-1$，即 $P_{n-1,m}$，$P_{n,m-1}$ 也在讨论范围内

考虑最后一张选票的所属，

如果最后一张投给了A，则前 $n-1+m$ 张选票符合 $P_{n-1,m}$ 的情形

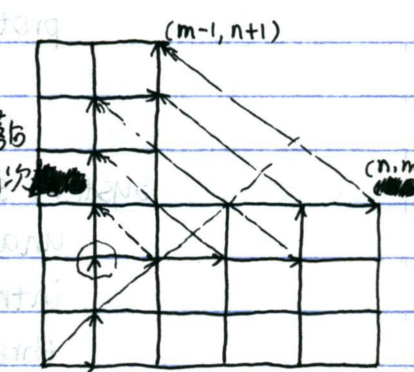
如果最后一张投给了B，则前 $n+m-1$ 张选票符合 $P_{n,m-1}$ 的情形

于是有 $P_{n,m}=P_{n-1,m}P(A)+P_{n,m-1}P(B)$

$$=\frac{n}{n+m}P_{n-1,m}+\frac{m}{n+m}P_{n,m-1}$$

于是可知 $P_{n,m}$ 符合初始条件 $P_{n,m}=\frac{1}{n+1}$，$n=m\ge 0$

和递推关系 $P_{n,m}=\frac{n}{n+m}P_{n-1,m}+\frac{m}{n+m}P_{n,m-1}$，$0\le m<n$



再考虑 $P_{n,m}$ 的闭公式，使用与卡塔兰数类似的方法

唱票顺序与从(0,0)到(n,m)的路径是一一对应的

考虑非法路径首次跨越虚线的时刻，此时对应于A的首次落后

将之后投给A的 $n-k$ 票记给B

而之后投给B的 $m-k-1$ 票记给A

则A总共得到 $m-1$ 票，而B总共得到 $n+1$ 票

于是可知A中途落后的唱票顺序与非法路径一一对应

即有 $\binom{n+m}{n+1}$ 种

于是有 $P_{n,m}=1-\binom{n+m}{n+1}/\binom{n+m}{n}=1-\frac{(n+m)!}{(n+1)!(m-1)!}\cdot\frac{n!\,m!}{(n+m)!}=\frac{n+1-m}{n+1}$

验证初始条件，当 $n=m$ 时，有 $P_{n,m}=\frac{n+1-n}{n+1}=\frac{1}{n+1}$

验证递推关系，当 $0<m<n$ 时，有

$$P_{n,m}=\frac{n}{n+m}P_{n-1,m}+\frac{m}{n+m}P_{n,m-1}=\frac{n}{n+m}\cdot\frac{n-m}{n}+\frac{m}{n+m}\cdot\frac{n+1-(m-1)}{n+1}$$

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

于是可知有闭公式 $P_{n,m}=\frac{n+1-m}{n+1}$，$0\le m\le n$

且卡塔兰数可以扩展至 $0\le m\le n$ 的情形 $C_{n,m}=\binom{n+m}{n}-\binom{n+m}{n+1}=\frac{n+1-m}{n+1}\binom{n+m}{n}$