1 概率统计基本概念

1.1 样本空间与随机事件

后面补上

1.2 概率公理化

定义1.1 (概率公理化定义)**.** *若随机试验E所对应的样本空间Ω中每一个事件A, 均赋予一实数$P(A)$, 如果满足以下条件:*

- *A1* 非负性: 对任一事件A, 有$P(A) \geq 0$;
- *A2* 规范性: $P(\Omega) = 1$;
- *A3* 可列可加性: 若$A_1, A_2, \ldots, A_n, \ldots$是两两不相容事件, 即对任意$i \neq j$有$A_i A_j = \emptyset$成立, 则有

$$P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots;$$

则称$P(A)$为事件A的概率.

根据概率的公理化定义, 可以推导概率的一些列重要性质.

性质1.1. $P(\emptyset) = 0$.

Proof. 首先有$\Omega = \Omega \cup \emptyset \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots$, 以及任意两两事件互不相容. 根据公理A2和公理A3得到

$$1 = P(\Omega) = P(\Omega) + P(\emptyset) + P(\emptyset) + \cdots = 1 + P(\emptyset) + P(\emptyset) + \cdots$$

于是有$P(\emptyset) = 0$成立. □

性质1.2 (有限可加性)**.** *若$A_1 A_2 \cdots A_n$是两两不相容事件, 则$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.*

Proof. 首先有$\cup_{i=1}^n A_i = \cup_{i=1}^n A_i \cup \cup_{i=n+1}^\infty \emptyset$, 以及任意两两事件互不相容. 根据公理A3有

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) + \sum_{i=n+1}^\infty P(\emptyset) = \sum_{i=1}^n P(A_i)$$

性质得证. □

性质1.3. *对任一事件A, 有$P(\overline{A}) = 1 - P(A)$.*

Proof. 因为$\Omega = \overline{A} \cup A$, 以及$A$与$\overline{A}$互不相容, 根据性质有限可加性有$P(\Omega) = 1 = P(A) + P(\overline{A})$. □

性质1.4. *对两事件A和B, 若$B \subset A$, 则有$P(A - B) = P(A) - P(B)$, 以及$P(B) \leq P(A)$.*

Proof. 若$B \subset A$, 有$A = B \cup (A - B)$, 由定义可知B与$A - B$互不相容, 于是有$P(A) = P(B) + P(A - B)$成立. 由公理A1可知$P(A - B) = P(A) - P(B) \geq 0$, 从而得到$P(A) \geq P(B)$. □

推广. *对两事件A和B, 有$P(A - B) = P(A) - P(AB) = P(A \cup B) - P(B)$.*

Proof. 首先有$A = (A - B) \cup (AB)$, 由定义可知$A - B$与AB互斥, 从而得到$P(A) = P(A - B) + P(AB)$. 另一方面, 有$A \cup B = (A - B) \cup B$, 以及$A - B$与$B$互斥, 于是有$P(A - B) = P(A \cup B) - P(B)$成立. □

性质1.5 (容斥原理 Inclusion-Exclusion Principle)**.** *对任意事件A和B, 有*

$$P(A \cup B) = P(A) + P(B) - P(AB).$$

Proof. 首先有$A \cup B = (A - B) \cup (AB) \cup (B - A)$, 由定义可知$A - B$, $B - A$, AB两两互不相容. 根据$P(A - B) = P(A) - P(AB)$和$P(B - A) = P(B) - P(AB)$, 有

$$P(A \cup B) = P(A - B) + P(B - A) + P(AB) = P(A) + P(B) - P(AB)$$

从而完成证明. □

推广. *对任意三个事件A, B, C, 有*

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC).$$

Proof. 两次利用容斥原理有

$$\begin{aligned} P(A \cup B \cup C) &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) + P(C) - P(AB) - P(AC \cup BC) \\ &= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \end{aligned}$$

从而完成证明. □

习题. *利用数学归纳法证明: 对于任意事件$A_1, A_2, \ldots, A_n$, 有*

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 \cdots A_n).$$

习题(Matching问题)**.** *有n对夫妻参加一次活动, 所有夫妻被随机两两分成n组, 每组1男1女, 问至少有一对夫妻被分到同一组的概率是多少?*

性质1.6 (Union Bound)**.** *对事件$A_1, A_2, \ldots A_n$, 有*

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leq P(A_1) + P(A_2) + \cdots + P(A_n).$$

Proof. 利用数学归纳法证明, 当$n = 2$时, 由容斥原理有

$$P(A \cup B) = P(A) + P(B) - P(AB) \leq P(A) + P(B). \tag{1}$$

假设当$n=k$时性质成立, 对$n=k+1$有

$$\begin{aligned} P(A_1\cup\cdots\cup A_{k+1}) &= P((A_1\cup\cdots\cup A_k)\cup A_{k+1}) \\ &\le P(A_1\cup\cdots\cup A_k)+P(A_{k+1}) \\ &\le P(A_1)+\cdots+P(A_k)+P(A_{k+1}) \end{aligned}$$

上式中第一个不等式成立是根据式(1), 而第二个不等式成立是根据归纳假设. □

推广(Bonferroni不等式)**.** 对事件$A_1, A_2, \ldots A_n$, 有

$$\begin{aligned} P(\cup_{i=1}^n A_i) &\le \sum_{i=1}^n P(A_i) \\ P(\cup_{i=1}^n A_i) &\ge \sum_{i=1}^n P(A_i)-\sum_{i<j}P(A_iA_j) \\ P(\cup_{i=1}^n A_i) &\le \sum_{i=1}^n P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k) \\ &\cdots \end{aligned}$$

例1.1. 设$P(A)=p, P(B)=q, P(AB)=r$, 用p,q,r表示事件的概率: *i)* $P(\overline{A}\cup\overline{B})$, *ii)* $P(\overline{A}B)$; *iii)* $P(\overline{A}\cup B)$; *iv)* $P(\overline{A}\cap\overline{B})$.

解. 对i), 根据定义可得

$$P(\overline{A}\cup\overline{B})=P(\overline{A})+P(\overline{B})-P(\overline{A}\cap\overline{B})=P(\overline{A\cap B})=1-r.$$

对于ii), 有$P(\overline{A}B)=P(B-A)=P(B)-P(AB)=q-r$. 对iii), 有

$$P(\overline{A}\cup B)=P(\overline{A})+P(B)-P(\overline{A}B)=1-p+q-(q-r)=1-p+r.$$

对iv), 根据i)可得$P(\overline{A}\cap\overline{B})=1-r-p-q$. □

习题. *1)* 已知$P(AB)=0$, 求证: $P(ABC)=0$.
2) 设$P(A)=P(B)=P(C)=\frac{1}{4}$, $P(AB)=0$, $P(AC)=P(BC)=\frac{1}{16}$, 问事件A,B,C至少有发生一个的概率.

1.3 古典概型

定义1.2 (古典概型)**.** 如果试验E满足

- 样本空间只包含有限个元素
- 每个基本事件发生的可能性相同

这类试验称为等可能概型, 又称古典概型.

值得注意的是: 古典概型要求每个基本事件发生的可行性相同. 下面看一个例子:

例1.2. *连续两次抛一枚均匀硬币, 观察事件: A) 两正面, B) 两反面, C)一正一反.*

根据古典概型可知$P(A) = P(B) = P(C) = 1/3$. 这个结论是不对的, 因为这三个事件发生的可能性不同. 正确的理解是事件$C = \{C_1, C_2\}$, 其中C_1表示先正后反的事件, C_2表示先反后正的事件, 从而有$P(A) = P(B) = P(C_1) = P(C_2) = 1/4$.

假设古典概型的样本空间$\Omega = \{w_1, w_2, \ldots, w_n\}$, 其中$w_i$为基本事件. 若事件$A$包含$k$个基本事件, 则

$$P(A) = k/n = |A|/|\Omega|,$$

这里$|A|$表示事件A包含的事件的个数, 由此可知古典概型的本质是计数(Counting). 下面介绍一些基本的计数原理.

设$\mathcal{A}_1$过程有n_1种方法, $\mathcal{A}_2$过程有n_2种方法.

- 加法原理: 若一项工作可用$\mathcal{A}_1$或$\mathcal{A}_2$来完成, 则完成该任务有$n_1 + n_2$种可能.
- 乘法原理: 若一项工作需分别通过$\mathcal{A}_1$和$\mathcal{A}_2$两过程, 则完成该任务需$n_1 \times n_2$种可能.

排列组合: 从N个元素中选出r个, 有

	有序	无序
有放回	N^r	$\binom{N+r-1}{r}$
无放回	$(N)_r$	$\binom{N}{r}$

其中$(N)_r = N(N-1)\cdots(N-r+1) = N!/r!$以及$\binom{N}{r} = N!/(r!(N-r)!)$.

解. 无放回且有序从N个元素中选出r个元素, 有$N(N-1)\ldots(N-r+1) = (N)_r = \binom{N}{r} \cdot r!$种方法;
无放回且无序从N个元素中选出r个元素, 有$\binom{N}{r}$种方法;
有放回且有序从N个元素中选出r个元素, 有N^r种方法;
有放回且无序从N个元素中选出r个元素, 将N个元素看作N个不同的盒子, r个无序的元素看作r个相同的球. 从N个不同元素取出r个无序的元素可看作将r个相同的球放入N 个顺序排序的盒子, 求不计放入球顺序的方法种类. 用'1'表示盒子, 用'$*$'表示球, 即

$$\underbrace{11\cdots1}_{n}\underbrace{**\cdots*}_{r}$$

将r个球放入n个盒子等价于

$$**1***1\cdots1$$

最后一个盒子固定, 两竖线之间表示球的个数, 所以共$\binom{N+r-1}{r}$种选法. □

下面看一些排列组合的例子.

例1.3. *将n只不同的球放入$N(N \geq n)$个不同的盒子, 求以下事件发生的概率:*

A: 恰有n个盒子且每盒一球;

B: 指定的n个盒子中各有一球;

C: 指定一盒子恰有m个球.

解. 将n只不同的球放入N个不同的盒子, 共有N^n种方法.
对事件A, 包含$N(N-1)\cdots(N-r+1)=(N)_n$种方法, 所以$P(A)=\frac{(N)_n}{N^n}=\frac{N!}{N^n r!}$.
对事件B, 包含$n!$种方法, 所以$P(B)=\frac{n!}{N^n}$.
对事件C, step 1: 指定盒子有m只球, 共有$\binom{n}{m}$种放法; step 2: 剩下$n-m$只球放入$N-1$个盒子, 共有$(N-1)^{n-m}$种放法. 所以$P(C)=\frac{\binom{n}{m}(N-1)^{n-m}}{N^n}$. □

例1.4 (抽签原理: 抽签的先后顺序不同是否会有优势?). 袋中有a只白球, b只红球(各不相同), 随机从中将球取出依次排成一列, 问第k次取出的球是红球的概率.

解. 用A表示第k次取到红球的事件, 则$P(A)=\frac{b(a+b-1)!}{(a+b)!}=\frac{b}{a+b}$(与$k$无关). 任何一人拿到红球的概率都是$\frac{b}{a+b}$, 抽签与先后顺序无关. □

习题. 在上例中, 如果白球完全相同, 红球完全相同, 如何求解.

例1.5. 有$k(k<365)$个人, 每个人的生日等可能地是365天的任意一天, 求至少两人生日相同的概率.

解. 令$A=\{$至少两人生日相同$\}$, $\overline{A}=\{k$个人生日均不同$\}$, 则$P(A)=1-P(\overline{A})=1-\frac{(365)_k}{(365)^k}$.
易知当$k=30$时, $P(A)=70.6\%$; $k=40$时, $P(A)=89.1\%$; $k=50$时, $P(A)=97\%$; $k=60$时, $P(A)=99.4\%$; $k=100$时, $P(A)=99.99\%$. □

例1.6. 设有N件产品, 其中有M件次品, 现从N件产品中任选n件, 求其中恰有k件次品的概率. 考虑两种情况: *i)* 不放回抽样, *ii)* 有放回抽样.

解. 对于不放回抽样, 从N件产品中任选n件, $\binom{N}{n}$种选法; n件产品恰有k件次品, $\binom{M}{k}\binom{N-M}{n-k}$种选法. 所以概率为$\frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$.
对于放回抽样, 每次抽到一件非次品的概率为$\frac{N-M}{N}$, 每次抽到一件次品的概率为$\frac{M}{N}$, 所以n件中恰有k件次品的概率是$\binom{n}{k}(\frac{M}{N})^k(\frac{N-M}{N})^{n-k}$. □

例1.7. 将n个男生和m个女人随机排成一列$(m<n)$, 任意两女生不相邻的概率是多少.

解. 先排n个男生, $n!$种排法; 后排m个女生, 要两两女生互不相邻, 有$(n+1)n\cdots(n-m+2)$种排法. 最后的概率为$\frac{n!(n+1)_m}{(m+n)!}=\frac{n!(n+1)!}{(m+n)!(n-m+1)!}$. □

习题. 若排列成一圈, 首尾相接, 则任意两女生不相邻的概率是多少.

例1.8. 从1至9个数中有放回取n个, 试求取出n个数的乘积被10整除的概率.

解. 令$A=\{$取出n个整数的乘积能被10整除$\}$, $B=\{$取出的n个数中有偶数$\}$, $C=\{$取出的n个数中至少有一个5$\}$. 显然, $A=BC$. 所以$P(A)=P(BC)=1-P(\overline{BC})=1-P(\overline{B}\cup\overline{C})=1-P(\overline{B})-P(\overline{C})+P(\overline{BC})=1-(\frac{5}{9})^n-(\frac{8}{9})^n+(\frac{4}{9})^n$. □

1.3.1 几何概型

传统的古典概型只讨论有限种等可能结果. 为将有限个推广到无限个, 引入概率分析的另一种方法: 几何方法.

定义1.3. 设有一可度量区域Ω(1, 2, 3维等), 在Ω内任意投点M, 投点在Ω内具有等可能性, 即落入Ω内的任意子区域A的可能性与A的测度成正比, 与A 的位置与形状无关. 这样的概率模型称之为**几何概型**. 事件A的概率为

$$P(A) = \frac{A\text{的测度}}{\Omega\text{的测度}}.$$

特点: 1) 无限性: 样本空间无限; 2) 等可能性: 每个样本点等可能发生.

例1.9. 甲乙二人约定在中午$12:00-13:00$到达某地约会, 到达时间完全随机, 且约定先到之人等另一人15分钟后离开. 求两人见面的概率.

解. 以x, y分别表示两人到达的时间, 因此样本空间$\Omega = \{(x,y)\big|0 \le x, y \le 60\}$.
设事件$A = \{(x,y)\big||x-y| \le 15\} = \{(x,y)\big|x-y \le 15\text{且}x-y \ge -15\}$, 于是有

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{60^2-45^2}{60^2} = 0.4375.$$

□

例1.10. 在区间$[0,1]$内随机取两个数, 求事件$A=${两数之积小于1/4}的概率.

解. 对任意$x, y \in [0,1]$, 定义样本空间$\Omega = \{(x,y)\big|0 \le x, y \le 1\}$. 事件$A$可以表示为$A = \{(x,y)\big|xy \le \frac{1}{4}\}$. 样本空间的测度$\mu(\Omega) = 1$, 而事件$A$的测度为

$$\mu(A) = \frac{1}{4} + \int_{\frac{1}{4}}^{1} \frac{1}{4x}\,dx = \frac{1}{4} + \frac{1}{4}\ln 2.$$

于是有$P(A) = \mu(A)/\mu(\Omega) = (1+\ln 2)/4$. □

另外一种用计算的方法求解概率: 统计模拟方法/™特卡洛采样(Monte Carlo). 用程序实现例 1.9:

$n_A \leftarrow 0$

For $i = 1 : N$

  $x \leftarrow Random(0, 60)$

  $y \leftarrow Random(0, 60)$

  If $|x-y| < 15$ Then

    $n_A \leftarrow n_A + 1$

  Endif

Endfor

Return n_A/N.