Ch 1-2 频率与概率

回顾前一次课

- 了解概率统计的发展历程
- **概率与统计**关系
- 随机现象：**二重性**
- 随机试验：三特点
- 样本空间、样本点
- 随机事件：基本事件、不可能事件、必然事件
- 事件关系：$\subset$ 、$=$、$\cup$、 $-$ 、$\cap$ 、$\bar{A}$ **互斥与对立事件**的关系
- 事件运算：幂等、交换、结合、**分配**、**对偶**

频率

随机事件在一次试验中可发生也可不发生, 通常关心随机事件发生可能性有多大, 为此引入频率，描述随机事件发生的频繁程度

定义 在相同的条件下，进行了n次试验，n次试验中事件A发生的次数为n_A，称为A发生的频数，事件A发生的频率为

$$f_n(A) = \frac{n_A}{n}$$

事件的频率在一定程度上反应了事件发生的可能性．频率的性质包括

- $0 \le f_n(A) \le 1$
- $f_n(\Omega) = 1$
- 若$A_1, A_2, \ldots A_k$两两互不相容，则$f_n(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} f_n(A_i)$

频率的稳定性

频率在试验中表现出**随机性**

在大量重复试验中，事件频率通常在一个常数p附近摆动，随着试验次数的增大摆幅越来越小，将这种规律称为 **频率的稳定性**

例如，历史上多人对重复投掷硬币的试验，下面是试验统计结果：

实验者	n	n_H	$f_n(H)$
德·摩根	2048	1061	0.5181
蒲 丰	4040	2048	0.5069
K·皮尔逊	12000	6019	0.5016
K·皮尔逊	24000	12012	0.5005

频率的稳定性-计算机模拟

利用计算机随机数对投币试验进行仿真，下图为实验结果

当试验次数不多时频率呈现波动性

当试验次数充分大时，频率具有稳定性

概率的统计定义

频率的稳定性即通常所说的统计规律性，是**随机事件本身所固有的客观属性**，可用于度量事件发生的可能性大小.

概率的统计定义 随机事件A在大量重复试验中发生的频率总是稳定地在一个常数p附近摆动，随着试验次数的增加而摆幅逐渐越小，则称常数p为事件A发生的**概率**，记为 $P(A)=p$.

概率的性质：

- $0 \le P(A) \le 1$
- $P(\Omega)=1$
- 若$A_1, A_2, \dots A_k$两两互不相容，则 $P(A_1 \cup A_2 \cup \cdots \cup A_k)=\sum_{i=1}^{k} P(A_i)$

概率与频率

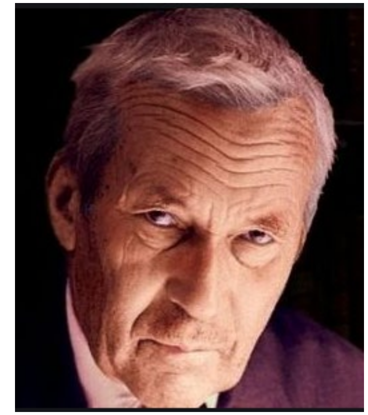
- 概率用于度量事件发生的可能性
- 频率在一定程度上反映了事件发生的可能性
- 概率是恒定的，而频率在试验中具有随机性
- 若试验次数足够多，频率与概率非常接近
- 概率可以通过频率来“测量”，频率是概率的一个近似

概率的统计定义存在数学上的不严谨性，在实际中几乎不可能每一个事件做大量重复的试验来计算频率，进而近似概率

受到频率的稳定性及其性质的启发给出严谨的概率公理化体系

概率的公理化定义

苏联数学家柯尔莫哥洛夫于1933年给出了概率的公理化定义，通过规定概率具备的基本性质来定义概率



概率的公理化定义 在随机试验的样本空间Ω上，对于每一个事件A 赋予一个实数，记为$P(A)$，若满足下列条件，称$P(A)$为事件A发生的**概率**：

- **非负性**：$P(A) \geq 0$
- **规范性**：$P(\Omega) = 1$
- **可列可加性**：若$A_1, A_2, \ldots$可列个两两互不相容的事件，则

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

简明扼要地刻画了概率的定义，为现代概率论奠定了基础，是概率论发展历史上的一个里程碑，从此被公认为数学的一个分支

概率的性质

- 对不可能事件∅有 $\mathbf{P(\emptyset) = 0}$

$\mathrm{A} = \emptyset \qquad \mathrm{P}(A) = 0$

$\mathrm{A} = \Omega \qquad \mathrm{P(A)} = 1$

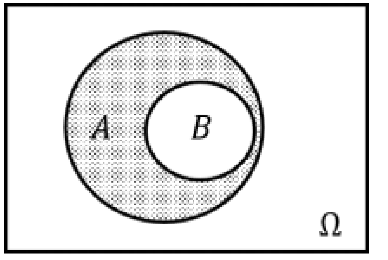
- **有限可加性** 若$A_1, A_2 \cdots A_n$是两两不相容事件, 则

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

- 对任意事件A有$\mathrm{P}(\overline{\mathrm{A}}) = 1 - \mathrm{P(A)}$

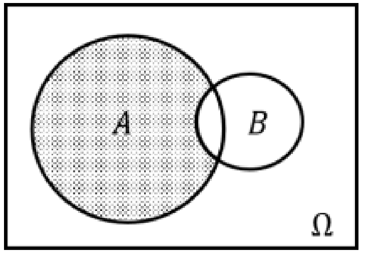
概率的性质

- 若$B \subset A$，则$P(A-B)=P(A)-P(B)$和$P(B) \le P(A)$



- 对任意事件A和B

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



容斥原理 (Inclusion-Exclusion Principle)

- 对任意随机事件A和B有

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

- 对三个随机事件A,B,C有

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

- 对事件$A_1,A_2,\cdots,A_n$有

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{i=1}^{n}P(A_i)-\sum_{i<j}P(A_iA_j)+\sum_{i<j<k}P(A_iA_jA_k)+\cdots+(-1)^{n-1}P(A_1\cdots A_n)$$

可进一步简化为

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{r=1}^{n}(-1)^{r+1}\sum_{i_1<\cdots<i_r}P(A_{i_1}\cdots A_{i_r})$$

不等式

Union bounds: 对事件 $A_1, A_2, \ldots A_n$ 有

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) \leqslant P(A_1) + P(A_2) + \cdots + P(A_n)$$

Bonferroni不等式：对事件 $A_1, A_2, \ldots A_n$ 有

$$\begin{aligned}
P\left(\bigcup_{i=1}^{n} A_i\right) &\leqslant \sum_{i=1}^{n} P(A_i); \\
P\left(\bigcup_{i=1}^{n} A_i\right) &\geqslant \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j); \\
P\left(\bigcup_{i=1}^{n} A_i\right) &\leqslant \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_iA_j) + \sum_{i<j<k} P(A_iA_jA_k);
\end{aligned}$$

可以依次类推

例子

设$P(A)=p,\ P(B)=q, P(AB)=r$, 用p, q, r分别表示下述事件的概率

1) $P(\bar{A}\cup\bar{B})$

2) $P(\bar{A}B)$

3) $P(\overline{\mathrm{A}}\cup B)$

4) $P(\overline{\mathrm{A}}\cap\bar{B})$

例子

设三个随机事件 A, B, C 满足

$$P(A) = P(B) = P(C) = 1/4,$$

$$P(AB) = 0, P(AC) = P(BC) = 1/16,$$

求事件 A, B, C 中至少有一个事件发生的概率.

Ch 1-3.1 古典概型

古典概型

概率论最早的研究对象，相对简单、在概率论中具有重要意义

如果试验E满足：

- 试验结果只有有限种可能，即 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$
- 每种结果发生的可能性相同，即 $P(\{\omega_i\}) = P(\{\omega_j\})\ (i \neq j)$

则称这样的试验为**古典概型**，又称**等可能概型**

根据上述定义可知：

$$P(\{\omega_i\}) = 1/n$$

若事件A包含k个基本事件，则事件A发生的概率为：

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

d'Alembert的推理

抛两枚硬币，观察向上的情况

有三种结果

A:两个正面　B:两个反面　C:一正一反

根据古典概率有：P(C) = 1/3

正确的解法

C为两部分：　C1:先正后反　C2:先反后正

四种结果：　A　B　C1　C2　等可能

正确的结果：$P(C) = P(C1) + P(C2) = 1/4 + 1/4 = 1/2$

基本计数原理

计数的两条基本原理

- **加法原理**: 若一项工作可以用两种不同的过程$\mathcal{A}_1$和$\mathcal{A}_2$完成, 且过程$\mathcal{A}_1$和$\mathcal{A}_2$分别有n_1和n_2种方法, 则完成该工作有$n_1 + n_2$种方法
- **乘法原理**: 若一项工作需要依次通过$\mathcal{A}_1$和$\mathcal{A}_2$两过程, 且过程$\mathcal{A}_1$和$\mathcal{A}_2$分别有n_1和n_2种方法, 则完成该工作有$n_1 \times n_2$种方法

上述两条原理可进一步推广到多个过程的情况

排列与组合

排列: 从n个不同的元素中无放回地取出r个元素进行排列，既要考虑取出的元素, 也要顾及其排列顺序, 有

$$(n)_r = n(n-1)\cdots(n-r+1)$$

种不同的排列。若$r = n$时称全排列, 有$n!$种

组合: 从n个不同的元素中无放回地取出r个元素, 取出的元素之间无顺序关系, 共有$\binom{n}{r}$种不同的取法，其中

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!}$$

称为**组合数**或**二项系数**

例题

将n个不同的球随机放入N $(N \geq n)$个不同的盒子中:

- 事件A表示恰有n个盒子且每盒一球
- 事件B表示指定的n个盒子中各有一球
- 事件C表示指定一盒子恰有m个球

求事件A, B, C发生的概率 (盒子容量不限, 放入同一盒子内的球无顺序区别)

生日问题

有k个人 ($k < 365$), 每个人的生日等可能地出现于365天中的任意一天, 求至少两人生日相同的概率

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{(365)_k}{365^k}$$

人数	概率
20	0.411
23	0.507
30	0.706
40	0.891
50	0.970
60	0.994
100	0.999999

超几何分布

设一批N件产品中有M件次品, 现从N件产品中不放回地任选n件, 求其中恰有k件次品的事件A的概率$P(A)$

超几何概率：

$$P(A) = \binom{M}{k}\binom{N-M}{n-k} \Big/ \binom{N}{n}$$

将上例中“无放回”修改为“有放回”，该问题如何求解？

抽签问题

袋中有a个不同的白球, b个不同的红球, 假设有k个人依次随机无放回地从袋中取一个球, 问第i个人 $(i \leq k)$ 取出红球的概率?

Matching问题

有n对夫妻参加一次聚会，现将所有参会人员任意分成n组，每组一男一女，问至少有一对夫妻被分到同一组的概率是多少?