第一章事件与概率

1.1 概率论发展简史

♠ French Society in the 1650's

- 赌博是一种流行时尚
- 法律没有禁止
- 当游戏越来越复杂，利益越来越大时，需要数学方法来计算获胜可能性

♠ The origin of the mathematical study of probability

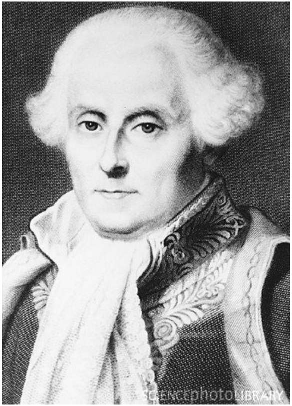
- 贵族**De Mere**在与一名宫廷卫士一次赌博时关于如何分赌本的问题发生了争执，于是请教他的好友数学家 **Blaise Pascal**.
- **Pascal** 与他的另一名好友数学家**Pierre Fermat**通信讨论了该问题，形成概率论中一个重要的基本概念—数学期望

♠ The first book on probability

- **Christiaan Huygens** 在 1657 年写了世界上第一本关于概率论的著作 **De ratiociniis in ludo aleae (“ On Reasoning in Games of Chance”)**, 中文译名 “论赌博中的计算”

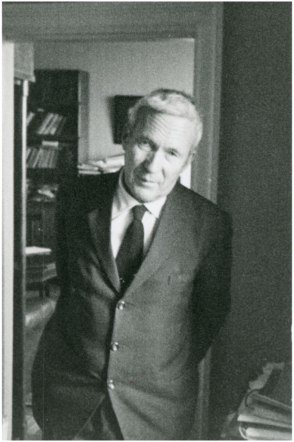
♠ From Games to Science

- **Pierre-Simon Laplace** 在他 1812 年的著作 **Theorie Analytique des Probabilities**中介绍了概率的数学理论及其科学应用.
- **Laplace**只考虑了古典概型，对一般的概率及其应用没有介绍.
- 到 1850 年, 许多数学家发现古典概型对一般场合不合理，开始尝试重新定义概率



♠ Axiomatic Development

- **Andrey Kolmogorov** 第一个在他 1933 年的著作 **Grundbegriffe der Wahrscheinlichkeit-srechnun**中严密的定义了概率.
- 类似于**Euclid**基于公理体系建立几何, 他从基本公理建立了概率理论, 从而使概率论称为一门严谨的数学分支
- 概率理论的现代研究和测度论非常紧密的结合在一起



概率论与数理统计的应用 概率统计理论与方法的应用几乎遍及所有科学技术领域、工农业生产和国民经济的各个部门中. 例如

- 气象、水文、地震预报、人口控制及预测都与**《概率论》**紧密相关
- 产品的抽样验收，新研制的药品能否在临床中应用，均要用到**《假设检验》**；
- 寻求最佳生产方案要进行**《实验设计》**和 **《数据处理》**；
- 电子系统的设计, 火箭卫星的研制及其发射都离不开**《可靠性估计》**
- 处理通信问题, 需要研究**《信息论》**；
- 研究经济数据等依时间观测数据时,**《时间序列分析》** 方法非常有用
- 研究化学反应的时变率，要以**《马尔可夫过程》** 来描述;

- 生物学中研究群体的增长问题时，提出了生灭型 **《随机模型》**，传染病流行问题要用到多变量非线性 **《生灭过程》**
- 许多服务系统，如电话通信、船舶装卸、机器维修、病人候诊、存货控制、水库调度、购物排队、红绿灯转换等，都可用一类概率模型来描述，其涉及到的知识就是 **《排队论》**
- 研究收入如何受个体教育程度, 行业, 性别等因素影响, 要用到 **《回归分析》**
- 研究多变量之间关系, 分类, 聚类等, 要用到 **《多元分析》**
- 研究寿命数据要用到 **《生存分析》**
- ...

1.2 概率论的几个基本概念

1.2.1 随机试验和随机事件

随机现象: 自然界中的一种客观现象, 当人们观测它时, 不能预先确定会出现哪种结果, 而仅仅知道是多种可能结果之一.

随机试验: 随机现象的实现和对它某个特征的观测.

- 随机试验中要求试验的结果至少 2 个
- 每次试验或观测得到其中的一个结果，在试验或观测之前不能预知是哪个结果发生
- 一般还要求试验在相同条件下能够重复

如观测把硬币抛 4 次后正面向上的次数; 观测某地的温度变化; 某电话总机单位时间内转接的电话次数.

基本事件：随机试验中的每个单一结果，它犹如分子中的原子，在化学反应中不能再分，所以有基本两字.

Definition

如把硬币抛 3 次后有 8 种可能结果：正正正、正正反、正反正、反正正、正反反、反正反、反反正、反反反. 这 8 种可能结果的每一个都是基本事件.

样本空间(Sample Space)： 随机试验中所有基本事件所构成的集合，通常用 Ω 或 S 表示. 样本空间中的元素，称为样本点，通常用 ω 等表示.

Definition

(1). 掷一枚骰子, 观察出现的点数. 则 $\Omega = \{1, 2, 3, 4, 5, 6\}$.

(2). 考察某一地区的年降雨量, 则 $\Omega = \{x | 0 \le x < T\}$, 这里 T 表示某个常数, 表示降雨量不会超过 T.

↑Example

↓Example

- 样本空间的元素应该是相互不同的, 根据试验的不同目的, 样本空间应该予以不同的选择.
- 但是总的原则是样本空间应该尽可能详细, 即尽可能包含所有可能的结果.

看下面的例子

(1). 将一枚硬币抛三次，考察正反面出现的情况;
(2). 将一枚硬币抛三次，考察正面出现的次数。

↑Example
↓Example

这两个试验的目的不同，因此样本空间的选取也不同。

随机事件：简称事件（Event），在随机试验中我们所关心的可能出现的各种结果，它由一个或若干个基本事件组成．

Definition

随机事件常用大写英文字母 A, B, C, D 等表示. 如果用语言表达, 则要用花括号括起来.

必然事件 (Ω)：在试验中一定会发生的事件；
不可能事件 (ϕ)：在试验中不可能发生的事件．

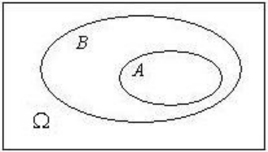
Definition

因此, 我们不严格的说**样本空间的子集称为随机事件**.

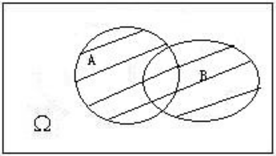
1.2.2 事件的关系及其运算

可以证明, 把样本空间中的基本事件与空间中的点相对应, 则事件与集合相对应, 因此事件运算与集合运算可以建立一一对应关系.

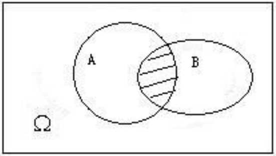
- 子事件 $A \subset B$: 事件 A 发生蕴含事件 B 一定发生, 则事件 A 称为事件 B 的子事件, 记为 $A \subset B$. 若 $A \subset B$, 且 $B \subset A$, 则称事件 A 与事件 B 相等, 记为 $A = B$.



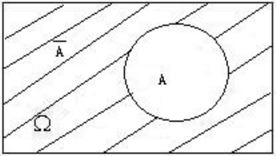
- 事件的和 $(A \cup B)$: 事件 A 和事件 B 中至少有一个发生的这一事件称为事件 A 和事件 B 的和, 记为 $A \cup B$.



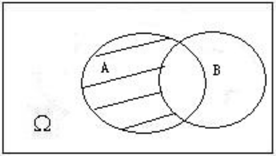
- 事件的积 $(A\cap B)$：事件 A 和事件 B 同时发生这一事件称为事件 A 和事件 B 的积, 记为 $A\cap B$. 如果 $A\cap B=\phi$, 则称 A 和 B 不相容或者互斥, 即事件 A 和 B 不能同时发生.



- 对立事件 A^c(或 $\bar{A}$): A 不发生这一事件称为事件 A 的对立事件(或余事件) .



- 事件 A 和事件 B 的差 $A-B$: 事件 A 发生而事件 B 不发生这一事件称为事件 A 和事件 B 的差, 记为 $A-B$, 或等价的, AB^c.



事件的运算

- 集合的运算法则适用于事件的运算

- De Morgan 对偶法则:

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \bar{A}_i$$

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \bar{A}_i$$

例：设 A, B, C 是三个事件，试表示下列事件

1. 事件 A, B 发生而 C 不发生;
2. 事件 A, B, C 不同时发生;
3. 事件 A, B, C 中至多有一个发生;
4. 事件 A, B, C 中至少发生两个;
5. 事件 A, B, C 中恰好发生两个;

↑Example
↓Example

1.2.3 概率的定义及性质

1. 概率的定义

什么叫概率? 直观地讲, 概率是随机事件发生可能性大小的数字表征, 其值习惯上用 0 和 1 之间的数表示, 换句话说, 概率是事件的函数. 如何求出事件 A 的概率 (记为 $P(A)$)?

(1) **古典概型**: 有两个条件,

> 第一 (有限性) 试验结果只有有限个 (记为 n) ,
> 第二 (等可能性) 每个基本事件发生的可能性相同.

为计算事件 A 的概率, 设 A 中包含 m 个基本事件, 则事件 A 的概率为

$$P(A) = \frac{m}{n} = \frac{|A|}{|\Omega|}$$

记号: 为方便起见, 以 $|B|$ 记事件 B 中基本事件的个数.

(2) 概率的统计定义

古典概型的两个条件往往不能满足，此时如何定义概率？常用的一种方法是把含有事件 A 的随机试验独立重复做 n 次 ($Bernouli$ 试验)，设事件 A 发生了 n_A 次，称比值 $\frac{n_A}{n}$ 为事件 A 发生的频率，当 n 越来越大时，频率会在某个值 p 附近波动，且波动越来越小，这个值 p 就定义为事件 A 的概率.

↑Example

抛硬币的试验

试验者	掷硬币的次数	正面出现的次数	频率
蒲丰	4040	2048	.5069
皮尔逊	12000	6019	.5016
皮尔逊	24000	12012	.5005

↓Example

从这个例子可以看出随着试验次数的增加，频率越来越接近 1/2.

(3) 主观概率

- 个人对某个结果发生可能性的一个判断, 如某人认为有 80% 的可能性房价暴跌. 另一人则认为仅有 20% 的可能性.
- 有相当的生活基础
- 在金融和管理等方面有大量应用
- 基于此的概率学派称为贝叶斯 (Bayes) 学派
- 但是当前用频率来定义概率的频率派仍是数理统计的主流. 焦点是频率派认为概率是客观存在, 不可能因人而异.

(4) **概率的公理化定义**: 对概率运算规定一些简单的基本法则,

称 $P(\cdot)$ 为一概率，如果

(i) 设 A 是随机事件，则 $0 \le P(A) \le 1$

(ii) 设 Ω 为必然事件，则 $P(\Omega) = 1$

(iii) 若事件 $A_1, A_2, \ldots$ 为两两不相容的事件序列，则

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

Definition

注: 事实上我们把可以定义 (计算) 概率的事件集合记为 $\mathcal{F}$, 其是样本空间的 $\sigma-$ 代数.

由概率的公理化定义, 我们可以得到

1. $P(\phi) = 0$

2. (有限可加性) 若 $A_k \in \mathcal{F}, k = 1, \cdots, n$ 且两两互斥，则

$$P(\sum_{k=1}^{n} A_k) = \sum_{k=1}^{n} P(A_k)$$

3. (可减性) 若 $A, B \in \mathcal{F}$ 且 $A \subset B$, 则 $P(B-A) = P(B)-P(A)$.

4. (单调性) 若 $A, B \in \mathcal{F}$ 且 $A \subset B$，则 $P(A) \leq P(B)$.

5. $P(\bar{A}) = 1 - P(A)$

6. (加法定理) 对任意的事件 $A_1, \cdots, A_n \in \mathcal{F}$，有

$$\begin{aligned} P(\sum_{k=1}^{n} A_k) &= \sum_{k=1}^{n} P(A_k) - \sum_{1 \leq i < j \leq n}^{n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n}^{n} P(A_i A_j A_k) \\ &\quad - \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n) \end{aligned}$$

7. (次可加性) 对任意的事件 $A_1, \cdots, A_n, \cdots \in \mathcal{F}$, 有 $P(\sum\limits_{n=1}^{\infty} A_n) \leq \sum\limits_{n=1}^{\infty} P(A_n)$.

8. (下连续性) 若 $A_n \in \mathcal{F}$ 且 $A_n \subset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\sum_{n=1}^{\infty} A_n) = \lim_{n} P(A_n)$$

9. (上连续性) 若 $A_n \in \mathcal{F}$ 且 $A_n \supset A_{n+1}, n = 1, 2, \cdots$, 则

$$P(\prod_{n=1}^{\infty} A_n) = \lim_n P(A_n)$$

(1992 年考研) 已知 $P(A)=P(B)=P(C)=1/4$, $P(AB)=0$, $P(AC)=P(BC)=1/6$, 求 $P(\bar{A}\bar{B}\bar{C})$.

↑Example

↓Example

解:

$$
\begin{aligned}
&P(\bar{A}\bar{B}\bar{C})=1-P(A\cup B\cup C)\\
&=1-[P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)]\\
&=1-3/4+2/6=7/12.
\end{aligned}
$$

从而

$$P(A\cup B\cup C)=5/12$$

但是又由 $P(A\cup B)=P(A)+P(B)-P(AB)=1/2$,

于是的矛盾

$$P(A\cup B)=1/2>5/12=P(A\cup B\cup C)$$

↑Example

求证对任意 n 个事件 $A_1, \cdots, A_n$ 有

$$P(\prod_{k=1}^{n} A_k) \geq \sum_{k=1}^{n} P(A_k) - n + 1$$

↓Example

作业.