Lecture 1

事件与概率

Y. Ruan
Department of Mathematics

Probability and Statistics
Beihang University

1.1 课程内容

- 预备知识: 微积分，多元微积分，线性代数
- 本课程是概率论与数理统计的入门课程，将向初学者介绍: 概率论，统计推断和随机过程;（1）概率论：概率空间，事件与概率，条件概率，随机变量，多元随机变量，随机变量的期望与方差，大数定理与中心极限定理等（2）统计推断：Bayes 估计，极大似然估计，样本估计量及其分布，假设检验（3）随机过程：平稳随机过程，平稳 Gaussian 过程, 遍历过程, Markov 链等
- 教学参考书目以及讲稿等课程相关内容可在

  https://scistats.github.io/找到

几个常见例子:

- 在一个计算机网络系统中，网络出现拥挤的概率有多大？
- 同一生产线生产的某产品，其合格率如何估计？
- 如何使用随机模型估算基于股票走势的金融产品的价格？
- 某一机构为客户提供服务，怎样分配资源才能让服务更有效？
- Google 搜索引擎是如何对页面进行排序的？
- 语音识别系统如何使用概率模型对音频序列进行识别？

1.2 集合

定义 1.2.1 **集合**是由一些对象组成的，这些对象称为集合的元素. 如果 S 是一个集合并且 x 是 S 的一个元素，我们写作 $x \in S$. 反之则写作 $x \notin S$. 空集记为 $\varnothing$.

例题 1.2.1 $S = \{x_1, x_2\}$ 是一个有限集. $S = \{x_1, x_2, x_3, \ldots\}$ 是一个可数集合. $S = \{x | 0 \leqslant x \leqslant 1\}$ 是区间 $[0, 1]$ 中所有实数组成的集合，这是一个不可数集合.

定义 1.2.2 如果集合 S 的元素同时也属于 T, 我们就称 S 是 T 的**子集**，并记为 $S \subset T$. 如果 $S \subset T$ 并且 $T \subset S$, 那么 $S = T$.

例题 1.2.2 包含所有元素的集合称是一个全集 Ω. 全集的定义取决于所考虑的问题. 当全集给定之后，所有集合都是全集的子集，特别地 $\varnothing$ 是全集的子集.

1.3　集合的运算

集合 S 的**余集**

$$S^c = \{x | x \notin S\} = \Omega \backslash S$$

并集

$$S \cup T = \{x | x \in S \text{ 或者} x \in T\}$$

$$\bigcup_{k=1}^{n} S_n = \{x | x \in S_k \text{ 对某个} k\}$$

交集

$$S \cap T = \{x | x \in S \text{ 并且} x \in T\}$$

$$\bigcap_{k=1}^{n} S_n = \{x | x \in S_k \text{ 对所有} k\}$$

定理 1.3.1 (De Morgan's Law) 对任意集合 S_1 , ..., S_n, 我们有

$$\left(\bigcup_{k=1}^{n} S_k\right)^c = \bigcap_{k=1}^{n} S_k^c, \left(\bigcap_{k=1}^{n} S_k\right)^c = \bigcup_{k=1}^{n} S_k^c$$

- 我们证明第一个等式. $x \in (\bigcup S_k)^c \Longleftrightarrow x \notin \bigcup S_k \Longleftrightarrow x \notin S_k$ 对所有 $k \Longleftrightarrow x \in S_k^c$ 对所有 $k \Longleftrightarrow x \in \bigcap S_k^c$. ///

1.4　排列组合知识点复习

- n 个不同对象的不同的排列方法总数为：$n!$
- 从 n 个不同对象中选择 k 个组成一个 $k-$ 序列，不同的 $k-$ 序列总数为：

$$\frac{n!}{(n-k)!}$$

- 从 n 个不同对象中任选 k 个（不计顺序）的方法总数为：

$$C_n^k = \frac{n!}{(n-k)!k!}$$

- 将 n 个不同对象划分为 r 组（不计顺序）, $n = n_1 + n_2 +$

$\ldots\ldots + n_r$, 那么不同的方法总数为：

$$C_n^{n_1,n_2,\ldots\ldots,n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

例题 1.4.1 设 k 为正整数. 定义在 $\mathbb{R}^n$ 中的光滑函数有多少个不同的 k 阶导数.

■ 设 $f(x_1, \ldots, x_n)$ 为光滑函数, $k = \alpha_1 + \alpha_2 \ldots \alpha_n$, $0 \leqslant \alpha_i \leqslant k$, 那么 k 阶导数可写为

$$\frac{\partial^k f(x_1, \ldots, x_n)}{\partial x_1^{\alpha_1} \ldots \partial x_n^{\alpha_n}}.$$

不同的求导方式可解释为: 有 k 个排列好的小球, 将 $n-1$ 个隔板放到每个小球的两侧 (分为 n 份), 不同的方法有多少种. 再进一步将其演变为: 现有 $n+k-1$ 个位置, 要将 k 个小球和

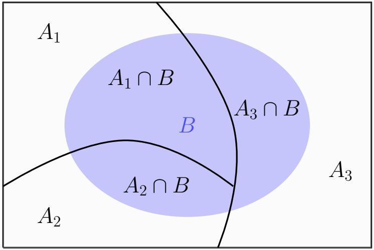
$n-1$ 个隔板放入这些位置的不同方法数

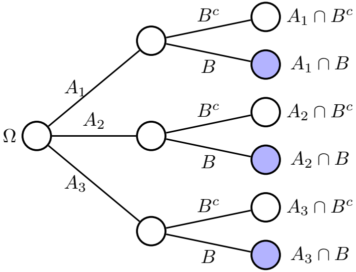
$$C_{n+k-1}^{k} = C_{n+k-1}^{n-1}.$$

1.5　概率模型

定义 1.5.1 **概率模型**由样本空间和概率组成. 样本空间是所有实验结果的集合. 样本空间的子集称为**事件**. **概率**则是集合上的函数，它赋予每一个事件一个非负数.

很多概率模型可以通过图形或是树形结构来表示:





样本空间

例题 1.5.1 将一个硬币投掷一次, 正反面概率分别为 $1/2$, 这里的样本空间是什么?

例题 1.5.2 在一次游戏中，玩家连续投掷一个硬币 10 次. 每出现一次正面，玩家赢得 1 元.

■ 在这个游戏中，影响到游戏结果的是正面出现的个数. 令 s_k $(k = 0, ..., 10)$ 表示在 10 次连续投掷中正面出现 k 次的事件. 那么我们可以定义样本空间为 $\Omega = \{s_0, s_1, ..., s_{10}\}$.

例题 1.5.3 在一次游戏中，玩家连续投掷一个硬币 10 次. 在第一个正面出现之前，每投一次硬币玩家赢得 1 元. 在第一个正面出现之后，直到第二个正面出现，每投一次硬币玩家赢得 2 元. 以此类推......

■ 在这个游戏中正面出现的次数以及何时出现都将影响到游戏结果. 因此合适的样本空间应该包括所有可能出现的10-投掷序列.

事件概率

例题 1.5.4 将一个均匀的硬币不断地投掷, 试求正面出现的概率.

■ 用 A 表示正面出现这一事件, 用 H_k 表示第一个正面出现时是在第 k 次投掷. 那么这些 H_k 表示的事件是相互排斥的, 因此

$$P(A) = \sum_{k=1}^{\infty} P(H_k)$$

$$= \sum_{k=1}^{\infty} \frac{1}{2^k} = \frac{1/2}{1-1/2} = 1.$$

这里的样本空间是什么呢?

例题 1.5.5 将一个不均匀的硬币不断地投掷, 每次投掷正面出现的概率为 $0 < p < 1$, 试求正面出现的概率.

- 用和上例同样的记号表示事件, 那么

$$P(A) = \sum_{k=1}^{\infty} P(H_k)$$
$$= \sum_{k=1}^{\infty} p(1-p)^{k-1} = p\frac{1}{1-(1-p)} = 1.$$

可见不论 p 有多小, 正面最终会出现.

例题 1.5.6 将一根长度为 L 的木棍分为三段，它们正好组成一个三角形的概率是多少?

■ 设分为三段后, 其中两段长度分别为 $0 < x, y < L$, 另一段长度为 $L - x - y > 0$. 满足这一条件的点对 (x, y) 在第一象限内表示为顶点为 $(0, L)$, $(0, 0)$, $(L, 0)$ 的三角形区域 Ω. 为能形成三角形, 三段的长度还应该满足

$$\begin{aligned} x + y &> L - x - y, \\ x + L - x - y &> y, \\ y + L - x - y &> x. \end{aligned}$$

化简得到

$$\begin{aligned} x + y &> \tfrac{L}{2}, \\ \tfrac{L}{2} &> y, \\ \tfrac{L}{2} &> x. \end{aligned}$$

这在坐标系中对应第一象限内顶点为 $(0, L/2)$, $(L/2, L/2)$, $(L/2, 0)$ 的三角形区域 E. 所关心的事件概率就是

$$\frac{Area\,(E)}{Area\,(\Omega)} = \frac{L^2/2}{(L/2)^2/2} = \frac{1}{4}.$$

■ 提示: 考虑一个高为 L 的等边三角形. 从该三角形的内部任意一点 p 向三边作垂线. 三条垂线段对应着一个分割木棍的方法. 注意到在图形Figure 1.1中 $L = a+b+c$. a, b, c 能组成一个三角形当且仅当 p 点位于绿色三角形中，绿色三角形是通过连接三边中点形成的. 所以所求的事件概率为 $1/4$.

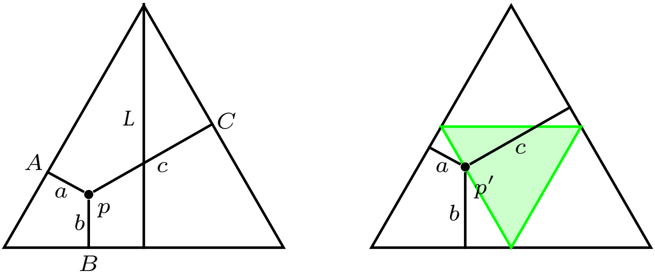


Figure 1.1: 三角形样本空间

例题 1.5.7 某班有 $N \geqslant 2$ 个学生, 现有 m 电影票, $1 \leqslant m < N$. 班长提出, 为了公平起见采用抓阄的方式来获得电影票: 即把电影票藏在纸团中, 同学们拍好队以后, 按次序抓阄. 请问这样的方式是否真的公平?

■ 事实上我们把抓阄的过程换个思路来解释, 首先让 N 个同学排好队, 将 m 张电影票放入 N 个纸团中, 每一个电影票的放入方法与已经排好队的同学对应之后就能得到一个抓阄结果.

为考察抓阄是否公平, 考虑排在第 k 位的同学, 他能抓到电影票就等价于: 将含有 m 张电影票的 N 个纸团进行排列, 在第 k 个纸团中有一张票. 而这一个事件的概率是

$$\frac{m\,(N-1)!}{N!} = \frac{m}{N}.$$

这与 k 无关, 因此抓阄是公平的. ///

概率公理

给定样本空间，**概率律**将赋予每一个事件 A 一个非负数 $P\,(A)$, 称之为事件 A 的概率. 概率律应满足一下公理

- $P(A) \geqslant 0$;
- $P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$ 如果 A_n 互不相交, 即 $\forall i \neq j$, $A_i \cap A_j = \varnothing$;
- $P(\Omega) = 1$.

定义 1.5.2 如果事件 A 满足 $P(A) = 1$, 那么我们称事件 A 是几乎必然发生的.

注意必然事件不一定是全集 Ω, 零概率事件也不一定是 $\varnothing$.

从概率公理可以推出一些简单的性质:

- $P(\varnothing) = 0$

- 令 $A_1=\Omega, A_n=\varnothing\ (n\geqslant 2)$. 运用概率公理得到

$$P\left(\bigcup_{n=1}^{\infty}A_n\right)=P(\Omega)=P(\Omega)+\sum_{n=2}^{\infty}P(\varnothing),$$

因此 $P(\varnothing)=0$.

- 假设 $A_n, n=1,\dots,N$ 为互不相交的集合, 即 $\forall i\neq j, A_i\cap A_j=\varnothing$. 那么

$$P\left(\bigcup_{n=1}^{N}A_n\right)=\sum_{n=1}^{N}P(A_n).$$

- 令 $A_n=\varnothing, n>N$. 对 $A_1,A_2,\dots,A_n,\dots$ 运用概率公理得

$$P\left(\bigcup_{n=1}^{N}A_n\right)=P\left(\bigcup_{n=1}^{\infty}A_n\right)=\sum_{n=1}^{\infty}P(A_n)=\sum_{n=1}^{N}P(A_n).$$

- 由于 $\Omega = A \cup A^c$, 总有 $P(A^c) = 1 - P(A)$
- 如果 $A \subset B$, 那么 $P(A) \leqslant P(B)$
- $P(B) = P(A \cap B) + P(A^c \cap B) \geqslant P(A \cap B) = P(A)$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- 注意到 $A \cup B$ 可以写成两个不相交集合的并集 $A \cup B = B \cup (A \cap B^c)$, 因此

$$\begin{aligned} P(A \cup B) &= P(B) + P(A \cap B^c) \\ &= P(B) + P(A) - P(A \cap B) \end{aligned}$$

- $P(A \cup B) \leqslant P(A) + P(B)$

事件概率的收敛性质

定理 1.5.1 任给递增事件序列 $\{A_n\}_{n=1}^{\infty}$, $A_1 \subset A_2 \subset \ldots$. 令 $A = \bigcup_{n=1}^{\infty} A_n$. 那么

$$P(A) = \lim_{n\to\infty} P(A_n)$$

- 令 $B_1 = A_1$, $B_2 = A_1^c \cap A_2$, $B_3 = A_2^c \cap A_3$, ... 那么 B_1, B_2, B_3, ... 互不相交, 并且

$$\bigcup_{i=1}^{n} B_i = A_n, \ \bigcup_{i=1}^{\infty} B_i = A$$

由概率的性质

$$P\left(A_n\right)=\sum_{i=1}^{n} P\left(B_i\right)$$

对 n 取极限

$$\begin{aligned}
\lim_{n\to\infty} P\left(A_n\right) &= \lim_{n\to\infty} \sum_{i=1}^{n} P\left(B_i\right) \\
&= \sum_{i=1}^{\infty} P\left(B_i\right) \\
&= P\left(\bigcup_{i=1}^{\infty} B_i\right) \\
&= P\left(A\right)
\end{aligned}$$

最后一个等号运用了概率公理.

定理 1.5.2 任给递减事件序列 $\{A_n\}_{n=1}^{\infty}$, $A_1 \supset A_2 \supset \ldots$. 令 $A = \bigcap_{n=1}^{\infty} A_n$. 那么

$$P(A) = \lim_{n \to \infty} P(A_n)$$

- 令 $B_1 = A_1^c, B_2 = A_2^c, B_3 = A_3^c, \ldots$ 那么 $B_1, B_2, B_3, \ldots$ 是递增事件序列, 并且

$$\bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i^c = A^c$$

对 $\{B_i\}_{i=1}^{\infty}$ 运用已证明的定理得到

$$P(A^c) = \lim_{n \to \infty} P(B_n)$$