《概率统计A》

主讲教师：任艳霞

http://www.math.pku.edu.cn/teachers/renyx/index.htm

请图灵班的同学不要选本门课!

- 上课时间: 周二1-2, 周四3-4(单)

  上课地点: 二教107

- 成绩评定:

  作业+ 期中+ 期末= $20 + 30 + 50$

- 考试内容: 以课堂讲授为主.

  每周二收发作业, 不接受迟交!

- 先修课: 高等数学(数学分析, 线性代数)
- 教材: 《概率与统计》, 第二版
  陈家鼎、郑忠国, 北京大学出版社.
- 进度安排
  第一部分 初等概率论(第一至四章) 约10次
  第二部分 随机过程(第五章) 约2次
  第三部分 数理统计(第六至十章) 约8次
- 期中考试暂定于第11次课, 内容: 前四章
- 期末考试内容: 后六章

第一章、随机事件与概率

- 第一次课

  §1.1 事件、概率

  §1.3 古典概模

- 第二次课

  §1.2 事件、概率的运算

  §1.4 概率的定义、性质

  §1.5 条件概率、独立性

- 第三次课

  §1.6 全概公式、逆概公式

  §1.7 独立试验序列

  §1.8 补充知识(自习)

§1.1. 随机事件及其概率

- 不确定性
- 例1.4. 投掷一枚硬币, 出现正面朝上.
- 随机试验(random trial), 不是"实验".
- **样本或样本点(sample)**: 每个试验结果称为一个样本点, ω.
- **样本空间**: 所有试验结果全体组成的集合Ω, (即为书上的S.)
- 事件(event): 部分试验结果, A, B,..., 是Ω 的子集.
- 概率(probability): 可能性, $P(A)$.

例1.4. 投掷一枚硬币(随机试验), 出现正面朝上(事件).
数学(模型)描述如下:

- 建模: H, 正面朝上(Head) ; T, 反面朝上(Tail),
- 两个样本(点), $\omega = H$ 或T,
- 样本空间, $\Omega = \{H, T\}$,
- 事件"出现正面朝上", $A = \{H\}$, 注: $A \neq H$.
- 概率, $P(A) = 1/2$ (公平硬币, 若无特别说明),
  $0 \leqslant P(A) \leqslant 1$ (一般情况).

概率$P(A)$ 的两个含义, 它们都不是概率的定义!

- 事件的频率(客观, §4.2 强大数律)

  例. 投n 次硬币, 出现m 次正面, 则$\frac{m}{n} \overset{n\text{很大}}{\approx} P(A)$.

  “频率”可作为概率的数值模拟(书P3 注1).

实验者	投掷次数	出现国徽朝上的次数	频率(μ/n)
德・摩根	2048	1061	0. 5181
蒲丰	4040	2048	0. 5069
皮尔逊	12000	6019	0. 5016
皮尔逊	24000	12012	0. 5005
维尼	30000	14994	0. 4998

- 事件的置信度(主观)

  信念, 把握, 经验, 预测.

§1.3 古典概率模型

古典概率模型: 在随机试验中, 总共有n 个试验结果, 每个结果等可能出现. 事件A 的概率为

$$P(A) := \frac{|A|}{|\Omega|}.$$

- 将试验结果(样本)分别记为 $\omega_1, \cdots, \omega_n$.
  令$A_i = \{\omega_i\}$, 它们被称为**基本事件**.
- 概率表达“权重”的分配, 但是不可以写$P(\omega_i)$.
- 不重不漏地数数, 排列, 组合, $C_n^m := \frac{n!}{m!(n-m)!}$.

1. 放回抽样

例1 盒中有m 个白球, n 个红球, 从中随机取出一个球, 记下颜色后放回, 然后再从盒中任取一个球. 求: (1) 抽到的都是白球的概率, (2) 抽到的两个球颜色不同的概率.

- 建模: 将白球编号: $1 \sim m$, 红球编号: $m+1 \sim m+n$.
  $\omega = (i,j) \neq \{i,j\}$: 依次抽到$i$ 号球, j 号球.
  $\Omega = \{(i,j) : 1 \leqslant i,j \leqslant m+n\}$, $|\Omega| = (m+n)^2$.
- $A =$“抽到的都是白球”$= \{(i,j) : 1 \leqslant i,j \leqslant m\}$, $|A| = m^2$.
  于是$P(A) = \frac{m^2}{(m+n)^2}$.
- $B =$“抽到的两个球的颜色不同”$= \{(i,j) : 1 \leqslant i \leqslant m < j \leqslant m+n$或$1 \leqslant j \leqslant m < i \leqslant m+n\}$,
  $|B| = 2mn$. 于是$P(B) = \frac{2mn}{(m+n)^2}$.

2. 不放回抽样

例3.3 (书) 有$N(\geqslant 2)$ 个阄, 其中$m(< N)$ 个内含"有"字. N 位同学依次任取一阄. 求: 第i 位同学取到"有"字阄的概率.

- 建模: 将阄编号: 含"有"字: $1 \sim m$, 剩余: $m+1 \sim N$.
  $\omega = (a_1, \cdots, a_N)$: 第$j$ 个人抽到编号为a_j 的阄.
  $\Omega = \left\{\omega : a_1, \cdots, a_N\text{是}1 \sim N\text{的排列}\right\}$, $|\Omega| = N!$.
- $A =$"第i 位同学取到'有'字阄"$= \{\omega : a_i \in \{1, \cdots, m\}\}$,
  $|A| = m(N-1)!$. 于是$P(A) = \frac{m}{N}$.
- 抓阄与顺序无关! 将同学编号$i_1, \cdots, i_N$.
  令$f : \omega \mapsto \tilde{\omega} = (a_{i_1}, \cdots, a_{i_N})$.
  $\tilde{\omega}$: 重排队列, 第i_1 号同学排第1, 第i_2 号同学排第2, $\cdots$.

例3.5 设有一批产品, 共100件, 其中5件次品. 现从中任取50件, 求: 恰好取到2件次品的概率.

- 建模: 将合格品编号: $1 \sim 95$, 次品编号: $96 \sim 100$.
  $\omega = \{a_1, \cdots, a_{50}\}$, $|\Omega| = C_{100}^{50}$.
  任取50件产品的一个试验结果记为D.
  $D \subset \{1, \cdots, 100\}$ 且$|D| = 50$. D 是样本, 不是事件.
- $A =$"恰好取到2件次品",

$$\begin{aligned} A &= \{D_1 \cup D_2 : \quad D_1 \subset \{1, \cdots, 95\}, |D_1| = 48; \\ &\qquad\qquad\qquad\quad D_2 \subset \{96, \cdots, 100\}, |D_2| = 2\}. \end{aligned}$$

  $|A| = C_{95}^{48} C_5^2$. $P(A) = C_{95}^{48} C_5^2 / C_{100}^{50}$.
- 自习: 例3.4, 3.6, 产品=阄, 次品(不合格)=有字的阄.

3.* 不可分辨物

- 由m 个1, n 个0组成的0-1字符串总共有C_{m+n}^{m} 个.

例2 在$0 \sim 99999$ 中随机抽取一个数, 求: 从左到右的数字单调上升的概率. $\Omega = 10^5$, $|A| =?$

- $|A| = C_{10-1+5}^{5}$:

$$22499 \mapsto ||| \circ \circ || \circ ||||| \circ \circ | \mapsto 11001101111100$$

$$03557 \mapsto \circ ||| \circ || \circ\circ || \circ || \mapsto 00111011001101 1$$

- $P(A) = C_{14}^{5}/10^5$.

- 将n个不同的元素分成有次序的k 组，不考虑每组中元素的次序，第i $(1 \leqslant i \leqslant k)$ 组恰有n_i 个元素的不同结果数为：

$$\frac{n!}{n_1!n_2!\cdots n_k!}$$

随机分组时, 得到的不同结果是等可能的.

- Jordan公式 设$A_1, A_2, \cdots, A_n$ 是n 个事件$(n \geqslant 2)$, 则

$$P(\bigcup_{i=1}^{n} A_i) = \sum_{k=1}^{n} (-1)^{k-1} \sum_{i_1<\cdots<i_k} P(A_{i1} A_{i_2} \cdots A_{i_k}).$$

- 自习: 例3.8(例2), 例3.9 (Jordan公式), 3.10 (匹配问题)

4.* 其他概率模型

- **离散型概率模型** (例4.2, 4.3)

  样本: $\omega_1, \omega_2, \cdots$, 分别分配权重$p_1, p_2, \cdots$;

  基本事件$A_i = \{\omega_i\}$ 的概率为p_i,

  事件A 的概率为$\sum_{i:\omega_i \in A} p_i$.

- **几何概型:**

  例: 某人随机于8点至9点之间到湖边, 逗留20分钟后离开.

  求: 9点前离开的概率.

  样本: $\Omega = [0, 1]$, $\omega \in [0, 1]$, 事件$A = [0, 2/3]$ 的概率为$2/3$.

例3* 20位男生与10位女生随机围成一个圈. 求: 每位女生的左右都是男生的概率.

- 以小红(女)为起点, 逆时针观察, 这是20位男生与9位女生的一个混排.
- 混排信息分解: 从29 个位置中挑9个给女生, 结果记为ω. $\Omega = \{\omega, \omega$是9个女生的位置$\}$, $|\Omega| = C_{29}^9$.
- $A =$女生的左右都是男生.
- 女生$=|$, 男生$= \circ$. A 发生当且仅当每个$|\,|$中至少有一个$\circ$, $|A| =$ 由9 个1 和10个0组成的0, 1 字符串的数目$= C_{19}^9$. $P(A) = \frac{|A|}{|\Omega|} = \frac{C_{19}^9}{C_{29}^9}$.