



# 概率论笔记

作者：肖程哲

时间：August 6, 2022



苟日新，日日新，又日新

## 目录

<b>第 1 章 概率基础</b>	<b>1</b>	3.3 估计与预测 . . . . .	29
1.1 概率空间 . . . . .	1	3.3.1 delta 法 . . . . .	29
1.1.1 随机事件 . . . . .	1	3.3.2 预测 . . . . .	29
1.1.2 概率空间 . . . . .	2	3.4 熵与信息 . . . . .	29
1.1.3 概率的性质 . . . . .	3	3.5 其他特征 . . . . .	29
1.2 古典概型与几何概型 . . . . .	4		
1.2.1 古典概型 . . . . .	4		
1.2.2 几何概型 . . . . .	5		
1.2.3 几何概型 . . . . .	5		
1.2.4 Bertrand 奇论 . . . . .	7		
1.3 条件概率 . . . . .	7		
1.4 独立性 . . . . .	10		
<b>第 2 章 随机变量</b>	<b>12</b>		
2.1 随机变量的分布 . . . . .	12		
2.2 多元随机变量 . . . . .	14		
2.2.1 边际分布 . . . . .	15		
2.2.2 条件分布 . . . . .	15		
2.2.3 独立 . . . . .	16		
2.3 随机变量的函数 . . . . .	16		
2.3.1 分布函数法 . . . . .	17		
2.3.2 Copula . . . . .	18		
2.3.3 概率密度函数法 . . . . .	19		
2.3.4 矩母函数法 . . . . .	20		
2.3.5 次序统计量 . . . . .	20		
<b>第 3 章 随机变量的数值特征</b>	<b>22</b>		
3.1 期望 . . . . .	22		
3.1.1 均值 . . . . .	22		
3.1.2 方差 . . . . .	23		
3.1.3 协方差 . . . . .	24		
3.1.4 条件期望 . . . . .	25		
3.2 矩母函数与特征函数 . . . . .	26		
3.2.1 矩 . . . . .	26		
3.2.2 矩母函数 . . . . .	27		
3.2.3 联合特征函数 . . . . .	28		
3.2.4 特征函数 . . . . .	28		
<b>第 4 章 常见分布</b>	<b>30</b>		
4.1 离散分布 . . . . .	30		
4.1.1 均匀分布 . . . . .	30		
4.1.2 伯努利分布 . . . . .	30		
4.1.3 二项分布 . . . . .	31		
4.1.4 几何分布 . . . . .	32		
4.1.5 负二项分布 . . . . .	33		
4.1.6 多项分布 . . . . .	34		
4.1.7 泊松分布 . . . . .	35		
4.1.8 超几何分布 . . . . .	37		
4.2 连续分布 . . . . .	38		
4.2.1 均匀分布 . . . . .	38		
4.2.2 指数分布 . . . . .	38		
4.2.3 伽马分布 . . . . .	40		
4.2.4 贝塔分布 . . . . .	41		
4.3 正态分布及其导出分布 . . . . .	43		
4.3.1 正态分布 . . . . .	43		
4.3.2 卡方分布 . . . . .	44		
4.3.3 F 分布 . . . . .	45		
4.3.4 t 分布 . . . . .	46		
4.3.5 柯西分布 . . . . .	47		
4.4 各分布间关系 . . . . .	49		
<b>第 5 章 概率极限</b>	<b>51</b>		
5.1 收敛 . . . . .	51		
5.2 大数定理 . . . . .	53		
5.3 中央极限定理 . . . . .	54		
<b>第 A 章 基本数学工具</b>	<b>56</b>		
A.1 集合论与测度论 . . . . .	56		
A.2 排列与组合 . . . . .	56		

# 第1章 概率基础

## 考试重点

□ 概率空间的定义

□ 古典概型与几何概型

□ 条件概率与独立

□ 伯努利试验

## 1.1 概率空间

### 1.1.1 随机事件

#### 定义 1.1 (样本空间)

随机试验可能出现的基本结果称为样本点 (sample point), 记为  $\omega$ 。样本的全体构成样本空间 (sample space), 记为  $\Omega = \{\omega\}$ 。若样本点有至多可数个, 则称为离散样本空间, 若不可数, 则称为连续样本空间。



#### 定义 1.2 (事件的古典定义)

样本点  $\omega$  的集合称为 (随机) 事件 (event)。



随机现象被抽象为集合。事件间的包含、相等、互不相容关系, 对应成集合间的包含、相等、不相交; 逻辑运算 (且、或、非等) 对应成集合论运算 (交、并、补等)。

**例题 1.1** 设  $A$ 、 $B$ 、 $C$  是某个随机现象的三个事件, 则

- 事件 “ $A$  与  $B$  发生,  $C$  不发生” 可表示为:  $AB\bar{C}$ .
- 事件 “ $A$ 、 $B$ 、 $C$  中至少有一个发生” 可表示为:  $A \cup B \cup C$ .
- 事件 “ $A$ 、 $B$ 、 $C$  中至少有两个发生” 可表示为:  $AB \cup AC \cup BC$ .
- 事件 “ $A$ 、 $B$ 、 $C$  中恰好有两个发生” 可表示为:  $AB\bar{C} \cup A\bar{B}C \cup \bar{A}BC$ .
- 事件 “ $A$ 、 $B$ 、 $C$  同时发生” 可表示为:  $ABC$ .
- 事件 “ $A$ 、 $B$ 、 $C$  都不发生” 可表示为:  $\bar{A} \cap \bar{B} \cap \bar{C}$ .
- 事件 “ $A$ 、 $B$ 、 $C$  不全发生” 可表示为:  $\bar{A} \cup \bar{B} \cup \bar{C}$ .

**性质** 集合的运算性质:

- 交换律

$$A \cup B = B \cup A, \quad AB = BA \quad (1.1)$$

- 结合律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC) \quad (1.2)$$

- 分配律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (1.3)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \quad (1.4)$$

- 对偶律 (De Morgan 公式)

$$\overline{A \cup B} = \bar{A} \cap \bar{B}, \quad (1.5)$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}. \quad (1.6)$$

**注** 并与补是集合中最基本的运算:

- 交的运算可通过并与对立来实现 (对偶律).

- 差的运算可通过对立与交来实现 ( $A - B = A\bar{B}$ ) .

### 1.1.2 概率空间

为方便概率的定义，避免不可测集的出现，并不把  $\Omega$  的一切子集作为事件。

#### 定义 1.3 (事件域)

事件构成的全体称为事件域  $\mathcal{F}$ ，是  $\Omega$  的子集族 (collection of subsets)，应满足  $\sigma$  代数 的要求：

- $\emptyset \in \mathcal{F}$ ，代表无事发生；
- $A \in \mathcal{F} \implies A^c \in \mathcal{F}$ ，即  $\mathcal{F}$  对补集运算 (逻辑上的非) 封闭；
- $A_1, \dots, A_n, \dots \in \mathcal{F} \implies \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ ，即  $\mathcal{F}$  对可数并运算封闭。



**注** 可数性是为了在数学上能够恰当地处理无穷的概念，术语中的  $\sigma$  指的就是可数并。由对偶原理可得  $\sigma$  域同时对可数并运算封闭。即  $\sigma$  域对逆，并，交，差的可数次运算封闭。

事件域根据问题的不同要求适当选取。在概率定义没有困难时，应尽量取得大，通常以  $\Omega$  的一切子集作为事件域。当  $\Omega$  给定后，若某些子集必须作为事件处理，能否找到包含他们的  $\sigma$  域？

#### 命题 1.1

若给定  $\Omega$  的一个非空集族  $\mathcal{G}$ ，必存在  $\Omega$  上唯一的  $\sigma$  域  $m(\mathcal{G})$ ，满足下列性质：

- 包含  $\mathcal{G}$
- 若其他  $\sigma$  域包含  $\mathcal{G}$ ，则必包含  $m(\mathcal{G})$

这个  $m(\mathcal{G})$  称为包含  $\mathcal{G}$  的最小  $\sigma$  域，或由  $\mathcal{G}$  扩张而成的  $\sigma$  域。



扩张，或者称为延拓，是数学中很重要的一个概念，大抵是将某映射的定义域适当扩大，不改变在初始定义域上的映射取值（注意值域可能是比较抽象的集合，配备了某些操作之后被称为空间），同时在扩大后的定义域上仍然保持某些优良的性质。与此相对的概念是限制，即关心局部上可能更加漂亮的性质，把初始的定义域适当缩小。

**证明** 由于  $\Sigma$  的一切子集构成的集类包含  $\mathcal{G}$ ，所以  $m$  存在。再取  $\Sigma$  上满足此条件的  $\sigma$  域之交作为  $m(\mathcal{G})$  即可。

特别地，实数集  $\mathbb{R}$  的子集族  $\{(-\infty, x] : x \in \mathbb{R}\}$  生成的  $\sigma$  代数  $\mathcal{B}_{\mathbb{R}}$  称为  $\mathbb{R}$  上的 Borel 代数。

#### 定义 1.4 (Borel 集)

设  $\mathbb{R}^1$  为全集，形为  $[a, b]$  构成的集类产生的  $\sigma$  域称为一维 Borel  $\sigma$  域，记为  $\mathcal{B}_1$ ，其中的元素称为一维 Borel 集



若  $x, y$  为任意实数，由于：

$$\{x\} = \bigcap_{n=1}^{\infty} \left[ x, x + \frac{1}{n} \right)$$

$$(x, y) = [x, y] - \{x\}$$

$$[x, y] = [x, y] + \{y\}$$

$$(x, y] = [x, y] + \{y\} - \{x\}$$

因此  $\mathcal{B}_1$  包含一切开区间，闭区间，单个实数，可列个实数，以及他们经可列次逆，并，交，差运算的集合。

#### 定义 1.5 (概率空间)

设  $\Omega$  为一个样本空间， $\mathcal{F}$  为定义于其上的一个事件域。定义在事件域(非样本空间) 上的集合函数  $P$ ：  
 $\mathcal{F} \rightarrow \mathbb{R}$  称为概率的条件是：

非负性  $\mathbb{P}(A) \geq 0, \forall A \in \mathcal{F}$

**规范性**  $\mathbb{P}(\Omega) = 1$ ; (如果没有这条就是一般的有限测度)

**可列可加性** 若  $A_1, \dots, A_n, \dots \in \mathcal{F}$  两两不交, 即  $A_i \cap A_j = \emptyset, \forall i \neq j$ , 则

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

并称  $(\Omega, \mathcal{F}, \mathbb{P})$  为一个概率空间 (probability space)



### 1.1.3 概率的性质

**性质** 概率的性质:

- $P(\Omega) = 1$ ;
- $P(A^C) = 1 - P(A)$ ;
- 若  $A \subseteq B$  则  $P(A) \leq P(B)$ ;

#### 推论 1.1 (加法公式)

基础形式:

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

一般形式:

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

特别地, 若事件出现个数相同时概率相等, 则可简化为:

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = nP_1 - \binom{n}{2}P_2 + \binom{n}{3}P_3 - \dots + (-1)^{n-1}P_n$$



显然, 可列可加性可以推出有限可加性. 但是一般来讲, 由有限可加性并不能推出可列可加性. 设  $A_i \in \mathcal{F}, i = 1, 2, \dots$  且两两互不相容, 若希望由有限可加性推出可列可加性, 则需要下式成立:

$$\lim_{n \rightarrow \infty} P\left(\sum_{i=1}^n A_i\right) = P\left(\lim_{n \rightarrow \infty} \sum_{i=1}^n A_i\right)$$

#### 定义 1.6

对于  $\mathcal{F}$  上的集合函数  $P$ , 若它对  $\mathcal{F}$  中任何一个单调不减的集序列  $\{S_n\}$  (即  $S_n \in \mathcal{F}, S_n \subseteq S_{n+1}$ ) 均满足:

$$\lim_{n \rightarrow \infty} P(S_n) = P\left(\lim_{n \rightarrow \infty} S_n\right)$$

则称它是下连续的.



故若令  $S_n = \sum_{i=1}^n A_i$ , 则可列可加性条件等价于有限可加性加下连续.

**例题 1.2** 口袋中有编号为  $1, 2, \dots, n$  的  $n$  个球, 从中有放回地任取  $m$  次, 求取出的  $m$  个球的最大号码为  $k$  的概率.

**解** 记事件  $A_k$  为“取出的  $m$  个球的最大号码为  $k$ ”. 如果直接考虑事件  $A_k$ , 则比较复杂, 因为“最大号码为  $k$ ”可以包括取到 1 次  $k$ 、取到 2 次  $k$ 、...、取到  $m$  次  $k$ .

为此我们记事件  $B_i$  为“取出的  $m$  个球的最大号码小于等于  $i$ ”,  $i = 1, 2, \dots, n$ . 则  $B$  发生只需每次从  $1, 2, \dots, i$  号球中取球即可. 所以由古典概率知

$$P(B_i) = \frac{i^m}{n^m}, \quad i = 1, 2, \dots, n.$$

又因为  $A_k = B_k - B_{k-1}$ , 且  $B_{k-1} \subset B_k$ , 由性质得

$$\begin{aligned} P(A_k) &= P(B_k - B_{k-1}) = P(B_k) - P(B_{k-1}) \\ &= \frac{k^m - (k-1)^m}{n^m}, \quad k = 1, 2, \dots, n. \end{aligned}$$

譬如,  $n = 6, m = 3$ , 可算得

$$P(A_4) = \frac{4^3 - 3^3}{6^3} = \frac{37}{216} = 0.1713.$$

其他的  $P(A_k)$  也都可算出, 现列表如下:

这相当于掷三颗骰子, 最大点数为 6 的概率是 0.4213, 而由

$$P(k \leq 3) = 0.0046 + 0.0324 + 0.0880 = 0.1250.$$

说明: 掷三颗骰子, 最大点数不超过 3 的概率仅为 0.1250.

一般而言, 求“至少有一个……”的概率时, 用对立事件公式去求较为方便. 但下面例 1.3 的配对问题却不能用对立事件去求解, 而一定要将事件“至少有个……”表示成事件的并, 然后用一般事件的加法公式去求解.

**例题 1.3 配对问题** 在一个有  $n$  个人参加的晚会上, 每个人带了一件礼物, 且假定各人带的礼物都不相同. 晚会期间各人从放在一起的  $n$  件礼物中随机抽取一件, 问至少有一个人自己抽到自己礼物的概率是多少?

解 以  $A_i$  记事件“第  $i$  个人自己抽到自己的礼物”,  $i = 1, \dots, n$ . 所求概率为  $P(A_1 \cup A_2 \cup \dots \cup A_n)$ . 因为

$$\begin{aligned} P(A_1) &= P(A_2) = \dots = P(A_n) = \frac{1}{n}, \\ P(A_1 A_2) &= P(A_1 A_3) = \dots = P(A_{n-1} A_n) = \frac{1}{n(n-1)}, \\ P(A_1 A_2 A_3) &= P(A_1 A_2 A_4) = \dots = P(A_{n-2} A_{n-1} A_n) = \frac{1}{n(n-1)(n-2)}, \\ &\dots \\ P(A_1 A_2 \dots A_n) &= \frac{1}{n!}. \end{aligned}$$

所以由概率的加法公式(??)得

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} - \frac{1}{4!} + \dots + (-1)^{n-1} \frac{1}{n!}.$$

譬如, 当  $n = 5$  时, 此概率为 0.6333; 当  $n \geq 10$  时, 此概率近似为  $1 - e^{-1} = 0.6321$ . 这表明: 即使参加晚会的人很多(譬如 100 人以上), 事件“至少有一个人自己抽到自己礼物”也不是必然事件.

## 1.2 古典概型与几何概型

### 1.2.1 古典概型

古典概型的基本思想:

**有限个样本点** 所涉及的随机现象只有有限个样本点, 譬如为  $n$  个, 且这些事件是两两互不相容的;

**等可能性** 每个样本点发生的可能性相等

#### 定义 1.7

若事件  $A$  含有  $k$  个样本点, 则事件  $A$  的概率为

$$P(A) = \frac{\text{事件 } A \text{ 所含样本点的个数}}{\Omega \text{ 中所有样本点的个数}} = \frac{k}{n}$$



**笔记** 古典概型的大部分问题都能形象化地用摸球模型来描述. 以后我们经常研究摸球模型, 意义即在于此.

**例题 1.4 生日问题**  $n$  个人的生日全不相同的概率  $p_n$  是多少?

**解** 把  $n$  个人看成是  $n$  个球, 将一年 365 天看成是  $N = 365$  个盒子, 则 “ $n$  个人的生日全不相同” 就相当于 “恰好有  $n$  ( $n \leq N$ ) 个盒子各有一球”, 所以  $n$  个人的生日全不相同的概率为:

$$P_n = \frac{365!}{365^n(365-n)!} = \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{n-1}{365}\right)$$

上式可用以下方法作近似计算:

1. 当  $n$  较小时, 原式右边中各因子的第二项之间的乘积  $\frac{i}{365} \times \frac{j}{365}$  都可以忽略, 于是有近似公式

$$p_n \approx 1 - \frac{1+2+\cdots+(n-1)}{365} = 1 - \frac{n(n-1)}{730}$$

2. 当  $n$  较大时, 因为对小的正数  $x$  有  $\ln(1-x) \approx -x$ , 所以由原式得

$$\ln p_n \approx \frac{1+2+\cdots+(n-1)}{365} = -\frac{n(n-1)}{730}$$

## 1.2.2 几何概型

古典概型的基本思想:

**有限个样本点** 所涉及的随机现象只有有限个样本点, 且这些事件两两互不相容的;

**等可能性** 每个样本点发生的可能性相等。

**例题 1.5 会商问题** 甲乙两人约定在下午 6 时到 7 时之间在某处会面, 并约定先到者应等候另一个人 20 min, 过时即可离去. 求两人能会面的概率.

**解** 以  $x$  和  $y$  分别表示甲、乙两人到达约会地点的时间 (以 min 为单位), 在平面上建立  $xOy$  直角坐标系 (见图 1.1).

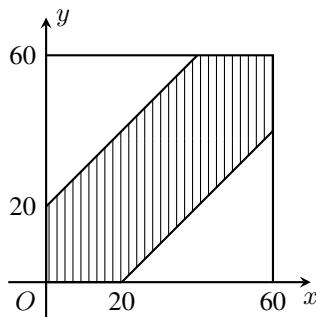


图 1.1: 会面问题中的  $\Omega$  与  $A$ .

因为甲、乙都是在 0 至 60 min 内等可能地到达, 所以由等可能性知这是一个几何概率问题.  $(x, y)$  的所有可能取值是边长为 60 的正方形, 其面积为  $S_\Omega = 60^2$ . 而事件  $A =$  “两人能够会面” 相当于:

$$|x - y| \leq 20,$$

即图中的阴影部分, 其面积为  $S_A = 60^2 - 40^2$ , 由 (??) 式知

$$P(A) = \frac{S_A}{S_\Omega} = \frac{60^2 - 40^2}{60^2} = 0.5556.$$

结果表明: 按此规则约会, 两人能会面的概率不超过 0.6. 若把约定时间改为在下午 6 时到 6 时 30 分, 其他不变, 则两人能会面的概率提高到 0.8889.

**例题 1.6 蒲丰投针问题** 平面上画有间隔为  $d$  ( $d > 0$ ) 的等距平行线, 向平面任意投掷一枚长为  $l$  ( $l < d$ ) 的针, 求针与任一平行线相交的概率.

**解** 以  $x$  表示针的中点与最近一条平行线的距离, 又以  $\varphi$  表示针与此直线间的交角, 见图 1.2. 易知样本空间  $\Omega$  满足

$$0 \leq x \leq d/2, \quad 0 \leq \varphi \leq \pi,$$

由这两式可以确定  $x - \varphi$  平面上的一个矩形  $\Omega$ , 这就是样本空间, 其面积为  $S_\Omega = d\pi/2$ . 这时为了针与平行线相交

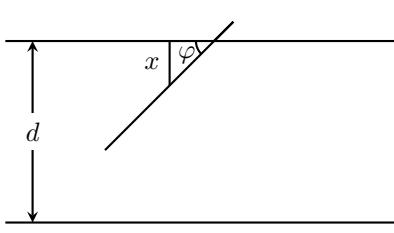
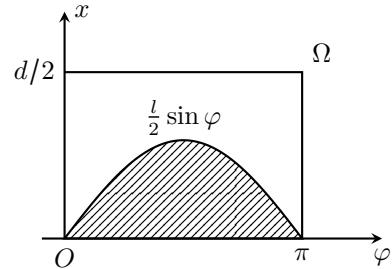


图 1.2: 蒲丰投针问题

图 1.3: 蒲丰投针问题中的  $\Omega$  和  $A$ 

(记为事件  $A$ ), 其充要条件是

$$x \leq \frac{l}{2} \sin \varphi.$$

由这个不等式表示的区域是图 1.3 中的阴影部分.

由于针是向平面任意投掷的, 所以由等可能性知这是一个几何概率问题. 由此得

$$P(A) = \frac{S_A}{S_\Omega} = \frac{\int_0^\pi \frac{l}{2} \sin \varphi \, d\varphi}{\frac{d}{2}\pi} = \frac{2l}{d\pi}.$$

如果  $l, d$  为已知, 则以  $\pi$  的值代入上式即可计算得  $P(A)$  之值. 反之, 如果已知  $P(A)$  的值, 则也可以利用上式去求  $\pi$ , 而关于  $P(A)$  的值, 可用从试验中获得的频率去近似它: 即投针  $N$  次, 其中针与平行线相交  $n$  次, 则频率  $n/N$  可作为  $P(A)$  的估计值, 于是由

$$\frac{n}{N} \approx P(A) = \frac{2l}{d\pi},$$

可得

$$\pi \approx \frac{2lN}{dn}.$$

历史上有一些学者曾亲自做过这个试验, 下表记录了他们的试验结果.

这是一个颇为奇妙的方法: 只要设计一个随机试验, 使一个事件的概率与某个未知数有关, 然后通过重复试验, 以频率估计概率, 即可求得未知数的近似解. 一般来说, 试验次数越多, 则求得的近似解就越精确. 随着电子计算机的出现, 入们便可利用计算机来大量重复地模拟所设计的随机试验. 这种方法得到了迅速的发展和广泛的应用. 入们称这种方法为**随机模拟法**, 也称为**蒙特卡罗 (Montecarlo) 法**.

**例题 1.7** 在长度为  $a$  的线段内任取两点将其分为三段, 求它们可以构成一个三角形的概率.

**解** 由于是将线段任意分成三段, 所以由等可能性知这是一个几何概率问题. 分别用  $x, y$  和  $a - x - y$  表示线段被分成的三段长度, 见图 1.4. 则显然应该有

$$0 < x < a; \quad 0 < y < a; \quad 0 < a - (x + y) < a.$$

第三个式子等价于:  $0 < x + y < a$ . 所以样本空间为 (见图 1.5)

$$\Omega = \{(x, y) : 0 < x < a, 0 < y < a, 0 < x + y < a\}.$$

$\Omega$  的面积为

$$S_\Omega = \frac{a^2}{2}.$$

又根据构成三角形的条件: 三角形中任意两边之和大于第三边, 得事件  $A$  所含样本点  $(x, y)$  必须满足:

$$0 < a - (x + y) < x + y,$$

$$0 < x < y + (a - x - y),$$

$$0 < y < x + (a - x - y).$$

整理得

$$\frac{a}{2} < x + y < a; \quad 0 < x < \frac{a}{2}; \quad 0 < y < \frac{a}{2}.$$

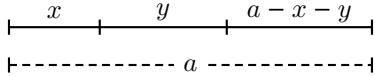
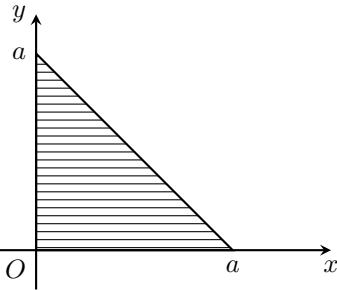
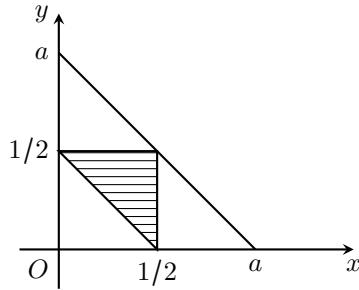
图 1.4: 长度为  $a$  的线段分成三段.图 1.5: 线段分成三段的样本空间  $\Omega$ .

图 1.6: 构成三角形的条件

所以事件  $A$  可用图 1.6 中的阴影部分表示。事件  $A$  的面积为

$$S_A = \frac{a^2}{8}.$$

由此得

$$P(A) = \frac{1}{4}.$$

### 1.2.3 Bertrand 奇论

## 1.3 条件概率

### 定义 1.8 (条件概率)

令  $A, B \in \mathcal{F}$ , 且  $P(B) > 0$  称

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

为基于于  $B$  的条件概率 (probability conditional on  $B$ ), 这仍然是一个概率测度.



### 定理 1.1 (乘法法则)

令  $A, B \in \mathcal{F}$ , 且  $P(B) > 0$ , 则

$$P(A \cap B) = P(A|B)P(B)$$

泛化后有:

$$P(A_1 \cap A_2 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\dots$$



**例题 1.8 罐子模型** 设罐中有  $b$  个黑球、 $r$  个红球, 每次随机取出一个球, 取出后将原球放回, 还加进  $c$  个同色球和  $d$  个异色球. 记  $B_i$  为“第  $i$  次取出的是黑球”,  $R_j$  为“第  $j$  次取出的是红球”.

若连续从罐中取出三个球, 其中有两个红球、一个黑球. 则由乘法公式我们可得

$$\begin{aligned} P(B_1R_2R_3) &= P(B_1)P(R_2|B_1)P(R_3|B_1R_2) \\ &= \frac{b}{b+r} \frac{r+d}{b+r+c+d} \frac{r+d+c}{b+r+2c+2d}, \\ P(R_1B_2R_3) &= P(R_1)P(B_2|R_1)P(R_3|R_1B_2) \\ &= \frac{r}{b+r} \frac{b+d}{b+r+c+d} \frac{r+d+c}{b+r+2c+2d}, \\ P(R_1R_2B_3) &= P(R_1)P(R_2|R_1)P(B_3|R_1R_2) \\ &= \frac{r}{b+r} \frac{r+c}{b+r+c+d} \frac{b+2d}{b+r+2c+2d}. \end{aligned}$$

以上概率与黑球在第几次被抽取有关.

罐子模型也称为波利亚 (Polya) 模型, 这个模型可以有各种变化, 具体见下:

- 当  $c = -1, d = 0$  时, 即为不返回抽样. 此时前次抽取结果会影响后次抽取结果. 但只要抽取的黑球与红球个数确定, 则概率不依赖其抽出球的次序, 都是一样的. 此例中有

$$\begin{aligned} P(B_1R_2R_3) &= P(R_1B_2R_3) = P(R_1R_2B_3) \\ &= \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}. \end{aligned}$$

例 ?? 可以归结为这种情况.

- 当  $c = 0, d = 0$  时, 即为返回抽样. 此时前次抽取结果不会影响后次抽取结果. 故上述三个概率相等, 且都等于

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br^2}{(b+r)^3}.$$

- 当  $c > 0, d = 0$  时, 称为传染病模型. 此时, 每次取出球后会增加下一次取到同色球的概率, 或换句话说, 每次发现一个传染病患者, 以后都会增加再传染的概率. 与前面两个一样, 以上三个概率都相等, 且都等于

$$\begin{aligned} P(B_1R_2R_3) &= P(R_1B_2R_3) = P(R_1R_2B_3) \\ &= \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}. \end{aligned}$$

从以上可以看出: 在罐子模型中只要  $d = 0$ , 则以上三个概率都相等. 即只要抽取的黑球与红球个数确定, 则概率不依赖其抽出球的次序, 都是一样的. 但当  $d > 0$  时, 就不同了.

- 当  $c = 0, d > 0$  时, 称为安全模型. 此模型可解释为: 每当事故发生了 (红球被取出), 安全工作就抓紧一些, 下次再发生事故的概率就会减少; 而当事故没有发生时 (黑球被取出), 安全工作就放松一些, 下次再发生事故的概率就会增大. 在这种场合, 上述三个概率分别为

$$\begin{aligned} P(B_1R_2R_3) &= \frac{b}{b+r} \frac{r+d}{b+r+d} \frac{r+d}{b+r+2d}, \\ P(R_1B_2R_3) &= \frac{r}{b+r} \frac{b+d}{b+r+d} \frac{r+d}{b+r+2d}, \\ P(R_1R_2B_3) &= \frac{r}{b+r} \frac{r}{b+r+d} \frac{b+2d}{b+r+2d}. \end{aligned}$$

### 定理 1.2 (全概率公式)

设  $B_1, B_2, \dots, B_n$  为样本空间  $\Omega$  的一个分割, 且互不相容, 即  $\bigcup_{i=1}^n B_i = \Omega, B_i B_j = \emptyset$  for  $i \neq j$ . 如果  $P(B_i) > 0, i = 1, 2, \dots, n$ , 则对任一事件  $A$  有

$$P(A) = \sum_{i=1}^n P(B_i)P(A|B_i). \quad (1.7)$$



笔记  $P(A|B_i)$  可视为事件  $A$  在  $B_i$  上的平均,  $P(B_i)$  则为其权重.

**定理 1.3 (Bayes 定理)**

设  $B_1, B_2, \dots, B_n$  为样本空间  $\Omega$  的一个分割, 且互不相容, 即  $\bigcup_{i=1}^n B_i = \Omega, B_i B_j = \emptyset$  for  $i \neq j$ . 如果  $P(A) > 0, P(B_i) > 0, i = 1, 2, \dots, n$ , 则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)} \quad (1.8)$$



**例题 1.9** 某地区居民的肝癌发病率为 0.0004, 现用甲胎蛋白法进行普查. 医学研究表明, 化验结果是存有错误的. 已知患有肝癌的人其化验结果 99% 呈阳性(有病), 而没患肝癌的人其化验结果 99.9% 呈阴性(无病). 现某人的检查结果呈阳性, 问他真的患肝癌的概率是多少?

**解** 记  $B$  为事件“被检查者患有肝癌”,  $A$  为事件“检查结果呈阳性”. 由题设知

$$\begin{aligned} P(B) &= 0.0004, & P(\bar{B}) &= 0.996, \\ P(A|B) &= 0.99, & P(A|\bar{B}) &= 0.001. \end{aligned}$$

我们现在的目的是求  $P(B|A)$ . 由贝叶斯公式得

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \\ &= \frac{0.0004 \times 0.99}{0.0004 \times 0.99 + 0.9996 \times 0.001} \\ &= 0.284. \end{aligned}$$

这表明, 在检查结果呈阳性的人中, 真患肝癌的人不到 30%, 这个结果可能会使人吃惊, 但仔细分析一下就可以理解了. 因为肝癌发病率很低, 在 10 000 个人中, 约有 4 人, 而约有 9996 个人不患肝癌. 对 10 000 个人用甲胎蛋白法进行检查, 按其错检的概率可知, 9996 个不患肝癌者中约有  $9996 \times 0.001 \approx 9.994$  个呈阳性. 另外 4 个真患肝癌者的检查报告中约有  $4 \times 0.99 \approx 3.96$  个呈阳性. 仅从 13.956 个呈阳性者中看, 真患肝癌的 3.96 人约占 28.4%.

**例题 1.10** 伊索寓言“孩子与狼”讲的是一个小孩每天到山上放羊, 山里有狼出没. 第一天, 他在山上喊: “狼来了! 狼来了!”, 山下的村民闻声便去打狼, 可到山上, 发现狼没有来; 第二天仍是如此; 第三天, 狼真的来了, 可无论小孩怎么喊叫, 也没有人来救他, 因为前两次他说了谎, 人们不再相信他了.

现在用贝叶斯公式来分析此寓言中村民对这个小孩的可信程度是如何下降的.

首先记事件  $A$  为“小孩说谎”, 记事件  $B$  为“小孩可信”. 不妨设村民过去对这个小孩的印象为

$$P(B) = 0.8, \quad P(\bar{B}) = 0.2. \quad (1.9)$$

我们现在用贝叶斯公式来求  $P(B|A)$ , 亦即这个小孩说了一次谎后, 村民对他可信程度的改变.

在贝叶斯公式中我们要用到概率  $P(A|B)$  和  $P(A|\bar{B})$ , 这两个概率的含义是: 前者为“可信”( $B$ ) 的孩子“说谎”( $A$ ) 的可能性, 后者为“不可信”( $\bar{B}$ ) 的孩子“说谎”( $A$ ) 的可能性. 在此不妨设

$$P(A|B) = 0.1, \quad P(A|\bar{B}) = 0.5.$$

第一次村民上山打狼, 发现狼没有来, 即小孩说了谎( $A$ ). 村民根据这个信息, 对这个小孩的可信程度改变为(用贝叶斯公式)

$$\begin{aligned} P(B|A) &= \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} \\ &= \frac{0.8 \times 0.1}{0.8 \times 0.1 + 0.2 \times 0.5} = 0.444. \end{aligned}$$

这表明村民上了一次当后, 对这个小孩的可信程度由原来的 0.8 调整为 0.444, 也就是(1.9)调整为

$$P(B) = 0.444, \quad P(\bar{B}) = 0.556. \quad (1.10)$$

在此基础上, 我们再一次用贝叶斯公式来计算  $P(B|A)$ , 亦即这个小孩第二次说谎后, 村民对他的可信程度改变为

$$P(B|A) = \frac{0.444 \times 0.1}{0.444 \times 0.1 + 0.556 \times 0.5} = 0.138.$$

这表明村民们经过两次上当，对这个小孩的可信程度已经从 0.8 下降到了 0.138，如此低的可信度，村民听到第三次呼叫时怎么再会上山打狼呢？

## 1.4 独立性

### 定义 1.9 (事件的独立性)

如果  $A, B \in \mathcal{F}$  满足

$$P(A \cap B) = P(A)P(B),$$

则称  $A$  与  $B$  独立 (independent)，记为  $A \perp\!\!\!\perp B$ .

对于事件集  $A_1, A_2, \dots, A_n$ ，若对于其中任意子集  $A_{i_1}, A_{i_2}, \dots, A_{i_n}$  有：

$$P(A_{i_1} \cap \dots \cap A_{i_n}) = P(A_{i_1}) \cdots P(A_{i_n})$$

则称此事件集相互独立 (mutually independent)



**笔记** 当  $P(A) > 0$  时，我们有  $P(B|A) = P(B) \iff B \perp\!\!\!\perp A$ ，由此可得到  $B$  独立于  $A$  的直观理解

**性质** 独立性是对称的，即  $A \perp\!\!\!\perp B \iff B \perp\!\!\!\perp A$ . 若两事件独立，则其补集也独立.

$$\begin{array}{c} A \leftrightarrow B \\ \nwarrow \nearrow \\ A^c \leftrightarrow B^c \end{array}$$

### 定义 1.10 (事件域的独立性)

若  $\mathcal{G} \subset \mathcal{F}$  与  $\mathcal{H} \subset \mathcal{F}$  满足

$$A \perp\!\!\!\perp B, \quad \forall A \in \mathcal{G}, B \in \mathcal{H},$$

则称  $\mathcal{G}$  与  $\mathcal{H}$  独立，记为  $\mathcal{G} \perp\!\!\!\perp \mathcal{H}$



测度论告诉我们一个重要结果：如果  $\mathcal{G}$  对交集运算封闭，那么成立  $\mathcal{G} \perp\!\!\!\perp \mathcal{H} \implies \sigma(\mathcal{G}) \perp\!\!\!\perp \mathcal{H}$

**例题 1.11** 有两名选手比赛射击，轮流对同一目标进行射击，甲命中目标的概率为  $\alpha$ ，乙命中目标的概率为  $\beta$ . 甲先射，谁先命中谁得胜. 问甲、乙两人获胜的概率各为多少？

**解** 记事件  $A_i$  为“第  $i$  次射击命中目标”， $i = 1, 2$ ，因为甲先射，所以事件“甲获胜”可以表示为

$$A_1 \cup \bar{A}_1 \bar{A}_2 A_3 \cup \bar{A}_1 \bar{A}_2 \bar{A}_3 \bar{A}_4 A_5 \cup \dots,$$

又因为各次射击是独立的，所以得

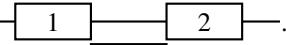
$$\begin{aligned} P(\text{甲获胜}) &= \alpha + (1 - \alpha)(1 - \beta)\alpha + (1 - \alpha)^2(1 - \beta)^2\alpha + \dots \\ &= \alpha \sum_{i=0}^{+\infty} (1 - \alpha)^i (1 - \beta)^i \\ &= \frac{\alpha}{1 - (1 - \alpha)(1 - \beta)}. \end{aligned}$$

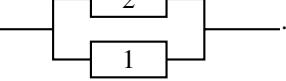
同理可得

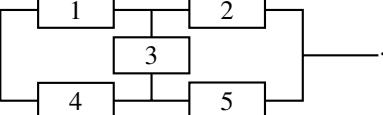
$$\begin{aligned} P(\text{乙获胜}) &= P(\bar{A}_1 A_2 \cup \bar{A}_1 \bar{A}_2 \bar{A}_3 A_4 \cup \dots) \\ &= (1 - \alpha)\beta + (1 - \alpha)(1 - \beta)(1 - \alpha)\beta + \dots \\ &= \beta(1 - \alpha) \sum_{i=0}^{+\infty} (1 - \alpha)^i (1 - \beta)^i \\ &= \frac{\beta(1 - \alpha)}{1 - (1 - \alpha)(1 - \beta)}. \end{aligned}$$

此题在等比级数求和时，应该有条件：公比  $|(1 - \alpha)(1 - \beta)| < 1$ . 这一点不难从题目的实际意义中得到。因为对本题而言， $\alpha, \beta$  取值为零或 1 均是无意义的。

**例题 1.12** 系统由多个元件组成，且所有元件都独立地工作。设每个元件正常工作的概率都为  $p = 0.9$ ，试求以下系统正常工作的概率。

(1) 串联系统  $S_1$ : 

(2) 并联系统  $S_2$ : 

(3) 5 个元件组成的桥式系统  $S_3$ : 

解 设  $S_i$  = “第  $i$  个系统正常工作”， $A_i$  = “第  $i$  个元件正常工作”。

(1) 对串联系统而言，“系统正常工作”相当于“所有元件正常工作”，即  $S_1 = A_1 A_2$ ，所以

$$P(S_1) = P(A_1 A_2) = P(A_1)P(A_2) = p^2 = 0.81.$$

这也可看出：两个正常工作概率为 0.9 的元件组成的串联系统，其系统正常工作的概率下降为 0.81。

(1) 对并联系统而言，“系统正常工作”相当于“至少一个元件正常工作”，即  $S_2 = A_1 \cup A_2$ ，所以

$$\begin{aligned} P(S_2)P(A_1 \cup A_2) &= P(A_1) + P(A_2) - P(A_1 A_2) \\ &= p + p - p^2 = 0.99. \end{aligned}$$

或

$$\begin{aligned} P(S_2) &= 1 - P(\bar{S}_2) = 1 - P(A_1 \bar{\cup} A_2) = 1 - P(\bar{A}_1 \cap \bar{A}_2) \\ &= 1 - P(\bar{A}_1)P(\bar{A}_2) = 1 - (1 - p)^2 = 0.99. \end{aligned}$$

这也可看出：两个正常工作概率为 0.9 的元件组成的并联系统，其系统正常工作的概率提高至 0.99。

(1) 在桥式系统中，第 3 个元件是关键，我们先用全概率公式得

$$P(S_3) = P(A_3)P(S_3|A_3) + P(\bar{A}_3)P(S_3|\bar{A}_3).$$

因为在“第 3 个元件正常工作”的条件下，系统成为先并后串系统（见图 1.7）。所以

$$\begin{aligned} P(S_3|A_3) &= P((A_1 \cup A_4)(A_2 \cup A_5)) = P(A_1 \cup A_4)P(A_2 \cup A_5) \\ &= [1 - (1 - p)^2]^2 = 0.9801. \end{aligned}$$

又因为在“第 3 个元件不正常工作”的条件下，系统成为先串后并系统（见图 1.8）。

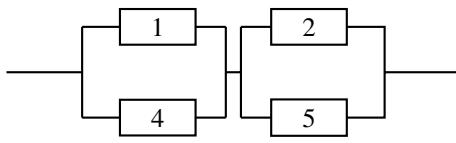


图 1.7: 先并后串系统

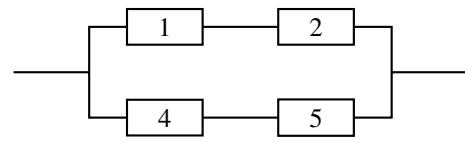


图 1.8: 先串后并系统

所以

$$P(S_3|\bar{A}_3) = P(A_1 A_2 \cup A_4 A_5) = 1 - (1 - p^2)^2 = 0.9639.$$

最后我们得

$$\begin{aligned} P(S_3) &= p[1 - (1 - p)^2]^2 + (1 - p)[1 - (1 - p^2)^2] \\ &= 0.9 \times 0.9801 + 0.1 \times 0.9639 = 0.9785. \end{aligned}$$