[entry]none/global/

第一章 随机事件及其概率

孙玲

网络空间安全学院

2022 年 9 月 5 日

lingsun@sdu.edu.cn

课件符号说明

概念/定义

- 概念/定义的内容.

性质/定理等

- 需要重点关注的性质/定理等内容.

统一书写格式

- 统一书写格式.

例题

- 例题.

补充内容

- 与课程相关的补充内容.

第一章 随机事件及其概率

确定性现象

- 一定条件下**必然发生**或**绝不可能发生**的现象.
  - 例 1: 在 1 个标准大气压下把水加热到 100℃ 时必然会沸腾.
  - 例 2: 在不受外力作用的条件下, 作等速直线运动的物体改变其等速直线运动状态是不可能的 (牛顿第一定律).

随机现象

- 在相同的条件下可能发生也可能不发生, 或者说, 可能出现这个结果, 也可能出现那个结果, 呈现出偶然性, 此类现象称为**随机现象**.
  - 例 1: 在落地之前, 抛一枚硬币究竟是正面朝上还是反面朝上是无法提前判定的.
  - 例 2: 同一门炮向同一目标发射多发同种炮弹, 弹落点也不一样, 一部分原因是炮弹制造时种种偶然因素对炮弹质量有影响. ◂ 前往弹坑原理 ◂ 前往统计规律性

弹坑原理

- 有战斗经验的老兵，在炮火连天的战场为了隐蔽自己，往往会跳进刚刚炸开的弹坑里，因为下一发炮弹，不容易落在同一点上，所以新弹坑是安全的.
  - 即使是用同样的发射器同样的动能在同一点将炮弹推出，还得考虑两次射击风速风向是不是完全一样 (完全一样可能性太小了)，只有完全一样的时候才是同一地点.
  - 两颗炮弹落到同一地点的几率非常小. 返回随机现象

统计规律性

在个别观察试验中呈现不确定的结果, 而在相同条件下, 大量重复试验, 试验结果呈现出的规律性称为**统计规律性**.

- 例 1: 大量重复抛掷同一枚硬币时正面朝上的次数约占抛掷总数的一半.
- 例 2: 衣服和用具总是在同样部位以相似的方式破损.
- 例 3: 下雨时地面各处总是差不多同时淋湿.

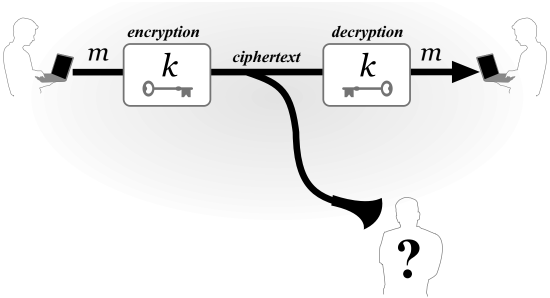
概率论与数理统计

- 概率论与数理统计是探究**随机现象统计规律性**的一门数学学科, 在自然科学、工程技术和社会科学的许多领域中有着重要的应用. 前往密码分析 前往机器学习 前往 §1.1

概率论在古典密码分析中的应用

密码学简介

- 从某种意义上来说，战争是促进科技发展最有效的催化剂.
- 纵观整个历史进程，每一次战争之后都会带来科技的快速飞跃，密码的产生和发展都与战争有着千丝万缕的关系.
- 自从有了战争，人类就产生了对安全通信的需求，密码也在这种背景下应运而生.



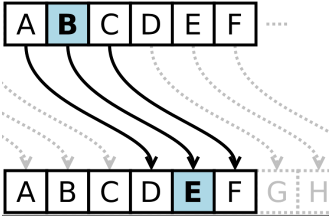
概率论在古典密码分析中的应用

密码学简介

- 密码的使用最早可追溯到公元前 400 多年，密码学的发展大致经历了四个阶段.
- **密码学发展的前夜时期（古代 ∼ 1948 年）**
  - 这一阶段密码的分析与设计没有严谨的数学理论支持.
  - 密码研究人员凭借自己的直觉和经验进行研究.
  - 这一阶段出现了一些比较经典的古典密码，如：凯撒（Caesar）密码和维吉尼亚（Vigenère）密码等。

凯撒密码 (Caesar Cipher)

- 密码学中，**凯撒密码**是一种最简单且最广为人知的加密技术.
- 它是一种替换加密的技术，明文中的所有字母都在字母表上向后/向前按照一个固定数目进行偏移后被替换成密文.
- 这个加密方法是以罗马共和时期凯撒的名字命名的，当年凯撒曾用此方法与其将军们进行联系.



- 例：APPLE $\rightarrow$ DSSOH.

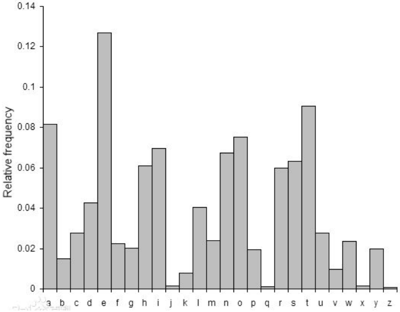
凯撒密码的破解

- 需要考虑两种情况:
  1. 攻击者知道/猜测密码中使用了某个简单的替换加密方式,但是不确定是凯撒密码.
  2. 攻击者知道/猜测使用了凯撒密码,但是不知道其偏移量.
- 对于第二种情况,解决方法更加简单. 由于使用凯撒密码进行加密的语言一般都是字母文字系统,因此密码中可能是使用的偏移量也是有限的,因而可以使用**穷尽搜索**的方法.
- 对于第一种情况,可以首先采用**频率分析法**来确定所使用的加密方式是否为凯撒密码.

概率论在古典密码分析中的应用

频率分析法

- 当密文长度足够大的情况下，可以先分析密文中每个字母出现的频率，然后将这一频率与正常情况下的该语言字母表中所有字母的出现频率做比较.
  - 例如在英语中，正常明文中字母 E 和 T 出现的频率特别高，而字母 Q 和 Z 出现的频率特别低.



频率分析法

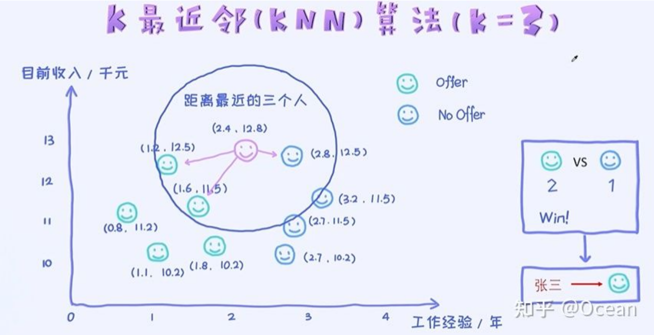
- 有时还可以将频率分析从字母推广到单词.
  - 例如英语中, 出现频率最高的单词是: the, of, and, a, to, in... 我们可以通过将最常见的单词的所有可能的 25 组密文, 编组成字典, 进行分析. 比如 QEB 可能是 the, MPQY 可能是单词 know.
- 频率分析也有其局限性, 它对于较短或故意省略元音字母或者其他缩写方式写成的明文加密出来的密文进行解密并不适用.

k-近邻算法 (k-Nearest Neighbor, KNN)

- 存在一个样本数据集合，也称作为**训练样本集** (Training Set)，并且样本集中每个数据都存在**标签** (Label)，即我们知道样本集中每一个数据与所属分类的对应关系.
- 输入没有标签的**新数据** (Test Set) 后，将新数据的每个特征与样本集中数据对应的特征进行比较，然后算法提取样本最相似数据 (最近邻) 的分类标签.
- 一般来说，我们只选择样本数据集中前 k 个最相似的数据，这就是 k-近邻算法中 k 的出处，通常 k 是不大于 20 的整数.
- 最后，选择 k 个最相似数据中**出现次数最多的分类**作为新数据的分类.

k-近邻算法示例

- 给定一些已经知道标签的样本（是否有 offer）.
- 判断张三是否有 offer.



本节主要内容

1. 随机试验与样本空间 ◂ 前往知识点
2. 随机事件 ◂ 前往知识点
3. 事件间的关系与运算 ◂ 前往知识点

◂ 前往 §1.2

1. 随机试验与样本空间

试验

- 为了确定随机现象的规律性，需要进行多次的试验、实验、调查或观察，我们把这些工作统称为**试验**.

随机试验的三个特征

1. 试验可在相同的条件下重复进行.
2. 试验的结果不止一个.
3. 每次试验之前，不能判定哪一个结果将会出现.

- 随机试验简称试验，用 E 表示.

随机试验的例子

- 掷一枚骰子，观察出现的点数.
- 观察日光灯的使用寿命.

1. 随机试验与样本空间

样本点 & 样本空间

- 试验 E 中的每一个可能的结果称为**基本事件/样本点**.
- 所有基本事件组成的集合称为试验 E 的**样本空间**, 记为 Ω.

样本空间 - 例 1

- 在投掷一枚硬币试验中, 共有两种可能的结果: "出现正面" 和 "出现反面".
- 这个试验的样本空间为 $\Omega = \{$"出现正面", "出现反面"$\}$.

样本空间 - 例 2

- 投掷一枚骰子, 观察其出现的点数, 所有可能出现的结果有 6 个, 分别是: 出现 1 点、出现 2 点、...、出现 6 点.
- 样本空间为

$$\Omega = \{\text{“出现 1 点”}, \text{“出现 2 点”}, \ldots, \text{“出现 6 点”}\}.$$

1. 随机试验与样本空间

样本点 & 样本空间

- 试验 E**基本事件/样本点**. E **样本空间**, Ω.

样本空间 - 例 3

- 在一批日光灯中任意抽取一只, 测试其寿命, 用 t (单位: 小时) 表示日光灯的使用寿命, 则 t 可取所有非负实数: $t \geqslant 0$.
- 样本空间为 $\Omega = \{t \mid t \geqslant 0\}$.

观测

- 例 1 和例 2 为离散型样本空间, 例 3 为连续型样本空间.
- 样本空间应根据随机试验的内容来确定.

返回

2. 随机事件

随机事件

- 随机试验 E 的样本空间 Ω 中的子集称为试验 E 的**随机事件**，简称**事件**，常用大写英文字母 $A, B, C, \ldots$ 表示.
- **基本事件**是最简单的随机事件.
- 一般的随机事件由若干个基本事件组成，称为**复合事件**.

随机事件 - 例子

- 投掷骰子，基本事件有：{出现 1 点}、{出现 2 点}、....
- 复合事件 {出现偶数点} 由基本事件 {出现 2 点}、{出现 4 点} 和 {出现 6 点} 组成.

统一书写格式

- 事件使用 $\{* * *\}$ 或 “$* * *$” 表示均可.

2. 随机事件

必然事件

- 由样本空间中的所有元素组成的集合，称为**必然事件**，用 Ω 来表示.

不可能事件

- 不含任何元素的空集合，称之为**不可能事件**，用 $\emptyset$ 来表示.

必然事件 & 不可能事件的例子

- "掷骰子，点数不超过 6"是必然事件.
- "掷骰子，点数大于 6"是不可能事件.

备注

- 严格来说，这两种事件不是随机事件，但为了后续讨论方便，仍把必然事件与不可能事件作为随机事件的一种特殊情况来统一处理.

3. 事件间的关系与运算

事件间的 6 种关系

1. 包含.
2. 事件的并.
3. 事件的交.
4. 事件的差.
5. 互不相容事件.
6. 对立事件.

讨论事件间关系的意义

- 讨论事件间关系有助于更好的理解事件的本质.

3. 事件间的关系与运算

关系 1: 包含

- 若事件 A 发生必然导致事件 B 发生, 则称事件 B 包含事件 A, 或称事件 A 包含于事件 B 中, 记 $B \supset A$ 或 $A \subset B$.

包含关系 - 性质

- 必然事件 Ω 包含任何事件 A, 事件 A 包含不可能事件 $\emptyset$, 即 $\Omega \supset A \supset \emptyset$.
- 若事件 A 包含事件 B, 且事件 B 也包含事件 A, 即 $A \supset B$ 且 $A \subset B$, 则称事件 A 与事件 B **相等**, 记 $A = B$.

3. 事件间的关系与运算

关系 2: 事件的并

- 若事件 A 与事件 B 至少有一个发生, 这样构成的事件称为事件 A 与事件 B 的**并事件**, 记 $A \cup B$.

事件的并 - 性质

- 事件 $A \cup B$ 通常包含三个部分: A 发生而 B 不发生, A 不发生而 B 发生, A 与 B 同时发生.
- n 个事件 $A_1, A_2, \ldots, A_n$ 的并事件为 $A_1 \cup A_2 \cup \cdots \cup A_n$.

事件的并 - 例子

- 10 件产品中有 3 件次品, 从中任取 2 件, 记
$A = \{$取到 1 件次品$\}$, $B = \{$取到 2 件次品$\}$, 则
$A \cup B = \{$至少取到 1 件次品$\}$.

3. 事件间的关系与运算

关系 3: 事件的交

- 由事件 A 与事件 B 同时发生而构成的事件称为事件 A 与事件 B 的**交事件**, 记 $A \cap B$.

事件的交 - 性质

- n 个事件 $A_1, A_2, ..., A_n$ 的交事件为 $A_1 \cap A_2 \cap \cdots \cap A_n$.

事件的交 - 例子

- 10 件产品中有 3 件次品, 从中任取 2 件, 记
  $A = \{$取到 1 件次品$\}$, $B = \{$取到 2 件次品$\}$,
  $C = \{$至多取到 2 件次品$\}$
- $A \cap B = \emptyset$.
- $A \cap C = A$.

3. 事件间的关系与运算

关系 4: 事件的差

- 事件 A 发生而事件 B 不发生, 这样构成的事件称为事件 A 与事件 B 的**差事件**, 记 $A-B$.

事件的差 - 例子

- 10 件产品中有 3 件次品, 从中任取 2 件, 记
  $A=\{$取到 1 件次品$\}$, $B=\{$取到 2 件次品$\}$,
  $C=\{$至多取到 2 件次品$\}$
- $C-B=\{$至多取到 1 件次品$\}$.

3. 事件间的关系与运算

关系 5: 互不相容事件

- 在一次试验中，若事件 A 与事件 B 不能同时发生，即 $A \cap B = \emptyset$，则称事件 A 与事件 B 为**互不相容事件/互斥事件**.

互不相容事件 - 性质

- n 个事件 $A_1, A_2, \ldots, A_n$，若他们之间两两互不相容，则称这 n 个事件是**互不相容**的. 此时，可记

$$A_1 \cup A_2 \cup \cdots \cup A_n = A_1 + A_2 + \cdots + A_n = \sum_{i=1}^{n} A_i.$$

互不相容事件 - 例子

- 投掷骰子，记 $A = \{出现\ 3\ 点\}$，记 $B = \{出现\ 4\ 点\}$，则事件 A 与事件 B 为互斥事件.

3. 事件间的关系与运算

关系 6: 对立事件

- 在一次试验中, 若事件 A 与事件 B 二者必有一个发生且仅有一个发生, 则称事件 A 与事件 B 为**对立事件/互逆事件**.
- 通常把 A 的对立事件记为 $\overline{A}$, 即 $B=\overline{A}$.
- $\overline{A}$ 称为 A 的逆事件.

对立事件 - 性质

- $A\cap\overline{A}=\emptyset$.
- $A\cup\overline{A}=\Omega$ (与互不相容事件区别).
- 若事件 A 与事件 B 对立, 则事件 A 与事件 B 互不相容, 反之不成立.

对立事件 - 例子

- 投掷硬币, 记 $A=\{出现图案\}$, $B=\{出现数字\}$, 则事件 A 与事件 B 为对立事件.

3. 事件间的关系与运算

概率论中事件间的关系 & 集合论中集合间的关系

- 形式上具有相似性.

符号	概率论	集合论
Ω	必然事件	全集
$\emptyset$	不可能事件	空集
A	事件	子集
$\overline{A}$	A 的对立事件	补集
$A \cup B$	事件的并	并集
$A \cap B$	事件的交	交集
$A - B$	事件的差	差集
$A \cap B = \emptyset$	互不相容事件	A 与 B 没有公共元素

3. 事件间的关系与运算

常用的运算律

- 集合论中常用的运算律同样适用于概率论中事件的运算.

1. 交换律: $A\cup B=B\cup A$, $A\cap B=B\cap A$.
2. 结合律: $(A\cup B)\cup C=A\cup(B\cup C)$, $(A\cap B)\cap C=A\cap(B\cap C)$.
3. 分配律: $A\cap(B\cup C)=(A\cap B)\cup(A\cap C)$,
   $A\cup(B\cap C)=(A\cup B)\cap(A\cup C)$.
4. 德摩根 (De Morgan) 定律: $\overline{A\cap B}=\overline{A}\cup\overline{B}$, $\overline{A\cup B}=\overline{A}\cap\overline{B}$.
   对于 n 个事件 A_1, A_2, ..., A_n, 有

$$\begin{aligned}\overline{A_1\cup A_2\cup\cdots\cup A_n} &= \overline{A_1}\cap\overline{A_2}\cap\cdots\cap\overline{A_n},\\ \overline{A_1\cap A_2\cap\cdots\cap A_n} &= \overline{A_1}\cup\overline{A_2}\cup\cdots\cup\overline{A_n}.\end{aligned}$$

常用的运算律 - 例子

- 教材 Page 6 例 1.1.5.

§1.2 随机事件的概率

随机现象数量指标

- 研究事件出现可能性大小对解决理论和实际问题非常重要.

概率满足的两个要求

- 事件的**概率**是刻画事件出现的可能性大小的一种数量指标,这个数量指标应满足以下两个要求:
  1. 它应是事件本身固有的,不随人的意志而改变的一种客观属性度量.
  2. 它必须符合一般常情,即事件发生可能性大的,它的值就大;事件发生可能性小的,它的值就小.

本节主要内容

1. 频率与概率
2. 古典概型
3. 几何概型
4. 概率的公理化定义

1. 频率与概率

定义 1.2.1

- 设在相同的条件下，进行了 n 次试验，在这 n 次试验中事件 A 出现了 m 次，则称

$$f_n(A)=\frac{m}{n}=\frac{\text{事件 } A \text{ 出现的次数}}{\text{试验的总次数}}$$

为随机事件 A 在 n 次试验中出现的**频率**，m 称为**频数**.

频率 - 性质

- 根据频率的定义，有 $0\leqslant f_n(A)\leqslant 1$，$f_n(\Omega)=1$，$f_n(\emptyset)=0$.
- 若 $A_1, A_2, \ldots, A_n$ 为两两互不相容事件，则

$$f_n(A_1\cup A_2\cup\cdots\cup A_n)=f_n(A_1)+f_n(A_2)+\cdots+f_n(A_n).$$

定义 1.2.1

- 设在相同的条件下，进行了 n 次试验，在这 n 次试验中事件 A 出现了 m 次，则称

$$f_n(A) = \frac{m}{n} = \frac{\text{事件 } A \text{ 出现的次数}}{\text{试验的总次数}}$$

为随机事件 A 在 n 次试验中出现的**频率**，m 称为**频数**.

频率 - 观测

- 事件的频率刻画了事件发生的频繁程度.
- 频率越大，事件 A 发生的越频繁，因而在一次试验中发生的可能性越大.
- 可否用频率来表示 A 在一次试验中发生的可能性的大小?

频率与概率的直观区别

- 随机事件的频率与我们进行的试验有关，而随机事件的概率则是完全客观的存在.
- 随机事件的频率 $f_n(A)$ 可以看作是概率的随机表现.

频率的稳定性

- 大量经验表明，当试验的次数足够大时，频率总是稳定于某一个常数附近，发生较大偏离的可能性很小，这一性质称为**频率的稳定性**.

频率的稳定性 - 例子

- 教材 Page 7 例 1.2.1 - 1.2.2.

1. 频率与概率

观测

- 随机事件在大量试验中存在着一种客观规律性，频率稳定性是这种规律性的表现.
- 频率稳定性由大量统计得来，因而也叫**统计规律性**.
- 可以用频率的稳定值作为事件发生可能性大小的度量指标.

定义 1.2.2 (概率的统计定义)

- 在大量重复试验中，若事件 A 发生的频率稳定地在某一常数 p 的附近摆动，则称该常数 p 为事件 A 发生的概率，记为 $P(A)$，且有 $P(A) = p$.

注意区分频率与概率

- 频率是变动的，概率为常数.
- 当试验次数足够多时，频率相对稳定，可把频率作为概率的近似值，即 $f_n(A) \approx P(A)$.
- 日常生活中遇到的产品合格率，彩票中奖率等一般指频率.

2. 古典概型

观测

- 如果使用概率的统计学定义，要确定一个事件的概率，需要进行大量重复试验.

古典概型的两个特点

- **有限性：** 试验样本空间中的样本点只有有限个，即基本事件的总数有限.
- **等可能性：** 每一次试验中，每个基本事件发生的可能性相等.

定义 1.2.3 (概率的古典定义)

- 设试验结果共有 n 个基本事件 $\omega_1, \omega_2, \ldots, \omega_n$，并且这些事件发生的可能性相等. 事件 A 由其中的 m 个基本事件组成，则事件 A 的概率为

$$P(A) = \frac{\text{事件 } A \text{ 包含的基本事件数}(m)}{\text{基本事件总数}(n)}.$$

古典概型 - 例子 (教材 Page 9: 例 1.2.3)

- 某城市电话号码升位后为八位数, 且第一位为 6 或 8. 求:
  (1) 从该城市中随机抽取一个电话号码为不重复的八位数的概率;
  (2) 从该城市中随机抽取一个电话号码后三位数都是 8 的概率.

2. 古典概型

古典概型 - 例子 (教材 Page 9: 例 1.2.3)

- 某城市电话号码升位后为八位数, 且第一位为 6 或 8. 求:
  (1) 从该城市中随机抽取一个电话号码为不重复的八位数的概率;
  (2) 从该城市中随机抽取一个电话号码后三位数都是 8 的概率.
- 古典概型解题思路:

Step 1 分析样本空间中样本点的个数.

- $n = 2 \times 10^7$.

Step 2 分析所考虑的随机事件中包含的样本点个数.

(1) $m = 2 \times 9 \times 8 \times 7 \times 6 \times 5 \times 4 \times 3$.
(2) $m = 2 \times 10^4$.

古典概型 - 例子 (教材 Page 9: 例 1.2.4)

- 设有一批产品共计 100 件, 其中有 80 件是正品, 20 件为次品, 现按两种方法抽取产品:
  - (a) 有放回抽样: 每次任意抽取一件, 经观察后放回, 再从中任取下一件.
  - (b) 不放回抽样: 每次任意抽取一件, 经观察后不放回, 从剩下的产品中任取一件.
- 分别按两种抽样方法, 求从这 100 件产品中任意抽取 3 件, 其中有 2 件次品的概率.

2. 古典概型

古典概型 - 例子 (教材 Page 9: 例 1.2.4)

- 设有一批产品共计 100 件，其中有 80 件是正品，20 件为次品，现按两种方法抽取产品：
  - (a) 有放回抽样：每次任意抽取一件，经观察后放回，再从中任取下一件.
  - (b) 不放回抽样：每次任意抽取一件，经观察后不放回，从剩下的产品中任取一件.
- 分别按两种抽样方法，求从这 100 件产品中任意抽取 3 件，其中有 2 件次品的概率.

(a) $n = 100^3$, $m = C_3^2 \times 20^2 \times 80$.

(b) $n = 100 \times 99 \times 98$, $m = C_3^2 \times 20 \times 19 \times 80$.

2. 古典概型

两种重要的抽样方式

1. **有放回抽样:** 每次任意抽取一件, 经观察后放回, 再从中任取下一件 (基本事件数量: n^m).
2. **不放回抽样:** 每次任意抽取一件, 经观察后不放回, 从剩下的产品中任取一件 (基本事件数量: $n!/(n-m)!$).

两种抽样方式之间的区别与联系

- 采用两种抽样方法求同一事件的概率很有可能是不同的.
- 当产品的总数较大, 而抽取的数量相对来说较小时, 在两种抽样方式下计算的概率值相差不大.
- 在实际问题中, 当产品的总数较大时, 可把不放回的抽样视为有放回抽样.

古典概型 - 例子

- 在 N 件产品中已知有 M 件次品，现从中随机抽取 n 件 (取后不放回). 问这 n 件中恰有 k 件次品的概率?

2. 古典概型

古典概型 - 例子

- 在 N 件产品中已知有 M 件次品，现从中随机抽取 n 件 (取后不放回). 问这 n 件中恰有 k 件次品的概率?
- 样本空间中基本事件的数量 $n = C_N^n$.
- 事件包含的基本事件数 $m = C_M^k \cdot C_{N-M}^{n-k}$.

2. 古典概型

古典概型 - 例子 (教材 Page 10: 例 1.2.5)

- 把 n 个不同的球以同样的概率随机放入 N $(N \geqslant n)$ 个盒子中, 假设每个盒子能容纳的球数没有限制, 试求下列各事件的概率:
  (1) 某指定的 n 个盒子中各有一个球;
  (2) 任意 n 个盒子中各有一个球;
  (3) 有一个盒子中恰有 m $(m \leqslant n)$ 个球.

古典概型 - 例子 (教材 Page 10: 例 1.2.5)

- 把 n 个不同的球以同样的概率随机放入 N ($N \geqslant n$) 个盒子中，假设每个盒子能容纳的球数没有限制，试求下列各事件的概率:
  (1) 某指定的 n 个盒子中各有一个球;
  (2) 任意 n 个盒子中各有一个球;
  (3) 有一个盒子中恰有 m ($m \leqslant n$) 个球.
- 样本空间中基本事件数 N^n.

(1) 事件所包含的基本事件数 $n!$.

(2) 事件所包含的基本事件数 $C_N^n \cdot n!$.

(3) 事件所包含的基本事件数 $C_n^m \cdot N \cdot (N-1)^{n-m}$.

例 1.2.5 的衍生 - 生日问题

- 求参加某次集会的 n 个人中至少有两人生日相同的概率 p_n.

例 1.2.5 的衍生 - 生日问题

- 求参加某次集会的 n 个人中至少有两人生日相同的概率 p_n.
- 把 n 个人看成 n 个小球, 把一年的 365 天看作盒子, $N = 365$.

$$\begin{aligned} p_n &= 1 - P(\{\text{所有人生日均不相同}\}) \\ &= 1 - \frac{C_N^n \cdot n!}{N^n}. \end{aligned}$$

  - 当 $n = 23$ 时, 同生日的概率已超过 $\frac{1}{2}$;
  - 当 $n = 40$ 时, 概率几乎达九成.
  - 当 $n = 60$ 时, 几乎可以肯定必有两人生日相同.

n	5	10	20	23	30	40	60
p_n	0.027	0.117	0.411	0.507	0.706	0.891	0.994

3. 几何概率

古典概型的局限性

- 试验的可能结果只能是有限个，并且要求等可能性.
- 推广第一步：样本空间中的样本点有无穷多个，但仍具备等可能性.

定义 1.2.4 (几何概率)

- 如果试验 E 的可能结果可以几何地表示为某区域 Ω 中的一个点 (区域 Ω 可以是一维的、二维的、三维的、...)，并且点落在 Ω 中某区域 A 的概率与 A 的测度 (长度，面积，体积等) 成正比，而与 A 的位置及形状无关，则随机点落在区域 A 的概率为

$$P(A) = \frac{A \text{ 的测度}}{\Omega \text{ 的测度}} = \frac{m(A)}{m(\Omega)},$$

称上式定义的概率为**几何概率**.

3. 几何概率

定义 1.2.4 (几何概率)

- 如果试验 E 的可能结果可以几何地表示为某区域 Ω 中的一个点 (区域 Ω 可以是一维的、二维的、三维的、...), 并且点落在 Ω 中某区域 A 的概率与 A 的测度 (长度, 面积, 体积等) 成正比, 而与 A 的位置及形状无关, 则随机点落在区域 A 的概率为

$$P(A)=\frac{A\text{ 的测度}}{\Omega\text{ 的测度}}=\frac{m(A)}{m(\Omega)},$$

称上式定义的概率为**几何概率**.

统一书写格式

- 统一用 $m(A)$ 表示事件 A 的测度.

3. 几何概率

几何概率 - 例子 (教材 Page 12: 例 1.2.7)

- 在线段 [0,3] 上任投一点, 求此点坐标小于 1 的概率.
- 几何概率解题思路:

Step 1 确定样本空间 Ω.

- $\Omega = [0, 3]$.

Step 2 确定事件 A 的范围.

- $A = [0, 1)$.

几何概率 - 例子 (教材 Page 12: 例 1.2.8)

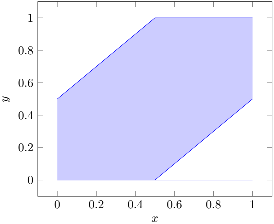
- 向区间 [0,1] 上任意投两点, 求此两点间距离小于 $\frac{1}{2}$ 的概率.

3. 几何概率

几何概率 - 例子 (教材 Page 12: 例 1.2.8)

- 向区间 [0,1] 上任意投两点，求此两点间距离小于 $\frac{1}{2}$ 的概率.

1. 确定样本空间 Ω.
   - x, y 表示两点坐标, $0 \leqslant x \leqslant 1$, $0 \leqslant y \leqslant 1$.
   - $\Omega = [0,1] \times [0,1]$.
2. 确定事件 A 的范围.
   - $|x - y| < \frac{1}{2}$.



几何概率 - 例子 (教材 Page 12: 例 1.2.9 蒲丰投针问题)

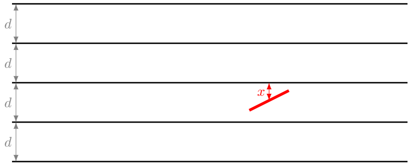
- 平面上画着一些平行线，它们之间的距离都等于 d，向此平面上任投一长度为 ℓ ($\ell < d$) 的针，求此针与任一平行线相交的概率.

几何概率 - 例子 (教材 Page 12: 例 1.2.9 蒲丰投针问题)

- 平面上画着一些平行线, 它们之间的距离都等于 d, 向此平面上任投一长度为 ℓ ($\ell < d$) 的针, 求此针与任一平行线相交的概率.

1. 确定样本空间 Ω.
   - x 表示针的中点到最近的一条平行线的距离, $0 \leqslant x \leqslant \frac{d}{2}$.
   - ϕ 表示针与平行线所成的夹角, $0 \leqslant \phi \leqslant \pi$.
   - $\Omega = \{(x, \phi) | 0 \leqslant x \leqslant \frac{d}{2}, 0 \leqslant \phi \leqslant \pi\}$.
2. 确定事件 A 的范围.
   - $x \leqslant \frac{\ell}{2} \cdot \sin\phi$.

- $p = \frac{2\ell}{\pi d}$.



3. 几何概率

蒲丰投针问题 - 引申

- 由于最后的答案与 π 有关，因此蒲丰设想利用这个试验来计算 π 的数值.
- 方法是投针 N 次，计算针与线相交的次数 n，再以频率值 $\frac{n}{N}$ 作为概率 p 带入 $p=\frac{2\ell}{\pi d}$，求得
$$\pi=\frac{2\ell N}{dn}.$$
- 虽然该方法耗力费时，而且很难达到当时用数学方法已算得小数点后一百多位精确数值的精度，然而在当时这个想法仍然是了不起的创意.
- 它提出的一种新的计算方案：建立一个概率模型，它与某些我们感兴趣的量有关，设计适当的随机试验，可以通过这个试验的结果来确定这些量.
- 随着计算机科学的发展，依靠这一思路已建立起一类新方法，即**随机模拟法**.

蒲丰投针问题 - 引申

- 下表给出一些基于这一想法的试验结果.

实验者	年份	针长	投掷次数	相交次数	π 的实验值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
De Morgan, C.	1860	1.0	600	382.5	3.137
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

4. 概率的公理化定义

几种概率定义的局限性

- 概率的统计定义
  - 要求进行大量统计试验.
  - 试验进行的次数 n 无法确定.
  - 无法证明增大试验的次数可以更逼近事件的概率.
- 古典定义和几何定义的假设前提是等可能性, 但不总成立.

4. 概率的公理化定义

- 苏联数学家柯尔莫哥洛夫在总结前人的研究基础上于 1933 年给出了概率的公理化定义.

概率的公理化定义

- 设随机试验 E 的样本空间为 Ω, 若对于 E 的每一个事件 A, 都有一个实数 $P(A)$ 与之对应, 并且满足下列三条公理, 那么称 $P(A)$ 为事件 A 的概率.
  - 公理 1 非负性: 对于任意一个事件 A, $0 \leqslant P(A) \leqslant 1$.
  - 公理 2 规范性: $P(\Omega) = 1$.
  - 公理 3 可列可加性: 对于两两互不相容事件 $A_1, A_2, \ldots, A_n, \ldots$, 有

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

4. 概率的公理化定义

概率的公理化定义

- 设随机试验 E 的样本空间为 Ω, 若对于 E 的每一个事件 A, 都有一个实数 $P(A)$ 与之对应, 并且满足下列三条公理, 那么称 $P(A)$ 为事件 A 的概率.
  - 公理 1 非负性: 对于任意一个事件 A, $0 \leqslant P(A) \leqslant 1$.
  - 公理 2 规范性: $P(\Omega) = 1$.
  - 公理 3 可列可加性: 对于两两互不相容事件 $A_1, A_2, \ldots, A_n, \ldots$, 有

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

- 公理的概率提取出了现实世界概率的共性, 并且在抽象之后, 概率可以**脱离现实问题背景**而变的更加威力强大.
- 数学家认为这些性质比概率的具体取值更能概括其本质.

本节主要内容

1. 概率的加法公式
2. 条件概率与事件的独立性.

1. 概率的加法公式

定理 1.3.1

- 若事件 A, B互不相容, 则

$$P(A+B) = P(A) + P(B).$$

证明

- 教材 Page 16.

概率的可加性一般形式

- 若事件 $A_1, A_2, \ldots, A_n$ 是两两互不相容事件, 则

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

统一书写格式

- 只有当事件互斥时, 事件的并才可以用 $+$ 来表示.
- 如果不确定两个事件是否互斥, 保险的方法仍然是使用 $\cup$.

1. 概率的加法公式

推论 0

- 不可能事件的概率为 0, 即 $P(\emptyset)=0$.

证明

- 因为

$$\Omega=\Omega+\emptyset$$

所以

$$P(\Omega)=P(\Omega)+P(\emptyset),$$

因此有 $P(\emptyset)=0$.

1. 概率的加法公式

推论 1

- 设 $\overline{A}$ 为 A 的对立事件，则

$$P(\overline{A}) = 1 - P(A),$$

证明

$$P(A) + P(\overline{A}) = P(A + \overline{A}) = P(\Omega) = 1$$

推论 1 - 应用情况

- 推论 1 为计算某些较复杂事件的概率提供了很大方便.

1. 概率的加法公式

定理应用实例 - 教材 Page 18: 例 1.3.1

- 设有产品 50 件, 其中有 3 件次品, 其余均为合格品, 今从中任意抽取 4 件, 求至少有一件次品的概率.

1. 概率的加法公式

定理应用实例 - 教材 Page 18: 例 1.3.1

- 设有产品 50 件，其中有 3 件次品，其余均为合格品，今从中任意抽取 4 件，求至少有一件次品的概率.
- 样本空间大小 C_{50}^4.
- 对立事件中基本事件数 C_{47}^4.

1. 概率的加法公式

推论 2

- 设事件 $A \supset B$, 则

$$P(A-B)=P(A)-P(B).$$

证明

- 因为 $A \supset B$, 所以 $A=(A-B)\cup B$.

推论 2.1 (概率的单调性)

- 若 $A \supset B$, 则 $P(A) \geqslant P(B)$.

1. 概率的加法公式

定理 1.3.2 (概率的一般加法公式)

- 若 A, B 为任意两事件，则

$$P(A\cup B)=P(A)+P(B)-P(A\cap B).$$

证明

- 因为 $A\cup B=A+(B-A\cap B)$，且有 $AB\subset B$.

推论 0 (布尔不等式)

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

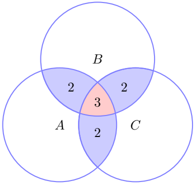
推论 1 (Bonferroni 不等式)

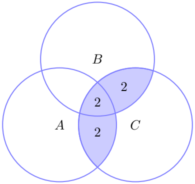
$$P(A\cap B)\geqslant P(A)+P(B)-1.$$

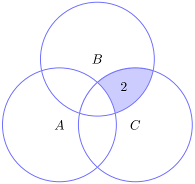
加法公式推广到 3 个事件

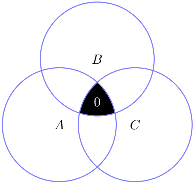
- 对于任意三个事件 A, B, C, 有

$$\begin{aligned} P(A\cup B\cup C) &= P(A)+P(B)+P(C) \\ &\quad -P(A\cap B)-P(A\cap C)-P(B\cap C) \\ &\quad +P(A\cap B\cap C). \end{aligned}$$









1. 概率的加法公式

加法公式推广到 3 个事件

- 对于任意三个事件 A, B, C, 有

$$\begin{aligned} P(A\cup B\cup C) &= P(A)+P(B)+P(C) \\ &\quad -P(A\cap B)-P(A\cap C)-P(B\cap C) \\ &\quad +P(A\cap B\cap C). \end{aligned}$$

一般加法公式

- 若 $A_1, A_2, \ldots, A_n$ 为 n 个事件, 则

$$\begin{aligned} P(A_1\cup A_2\cup\cdots\cup A_n) &= \sum_{i=1,2,\cdots,n} P(A_i) - \sum_{\substack{i<j \\ i,j=1,2,\cdots,n}} P(A_i\cap A_j) \\ &\quad + \sum_{\substack{i<j<k \\ i,j,k=1,2,\cdots,n}} P(A_i\cap A_j\cap A_k) - \cdots \\ &\quad +(-1)^{n-1}P(A_1\cap A_2\cap\cdots\cap A_n). \end{aligned}$$

1. 概率的加法公式

概率计算的公式法

- 利用上述公式来作概率计算常能使解题思路清晰，计算便捷.

概率计算 - 例子 - 教材 Page 18: 例 1.3.2

- 袋中装有红、白、黑球各一个，每次从袋中任取一个球，记录其颜色以后再放回袋中，这样连取 3 次 (有放回地抽取). 求 3 次都没有取到红球或 3 次都没有取到白球的概率.

1. 概率的加法公式

概率计算的公式法

- 利用上述公式来作概率计算常能使解题思路清晰，计算便捷.

概率计算 - 例子 - 教材 Page 18: 例 1.3.2

- 袋中装有红、白、黑球各一个，每次从袋中任取一个球，记录其颜色以后再放回袋中，这样连取 3 次 (有放回地抽取). 求 3 次都没有取到红球或 3 次都没有取到白球的概率.
- $A=\{3$ 次都没有取到红球$\}$, $B=\{3$ 次都没有取到白球$\}$.
- $P(A\cup B)=P(A)+P(B)-P(A\cap B)$.
- $P(A)=P(B)=\frac{2^3}{3^3}$.
- $P(A\cap B)=\frac{1^3}{3^3}$.

2. 条件概率与事件的独立性

条件概率 - 使用情境

- **已知某一事件 B 已经发生**, 要求另一事件 A 发生的概率.

条件概率 - 例子

- 考虑有两个孩子的家庭, 假定男女出生率一样, 则两个孩子 (依大小排序) 的性别为: (男, 男), (男, 女), (女, 男), (女, 女). 四种情况可能性是一样的.
- 记 $A = \{$有一个男孩, 有一个女孩$\}$, 则 $P(A) = \frac{1}{2}$.
- 如果预先知道至少有一个女孩, 那么事件 A 的概率为 $\frac{2}{3}$.
- 两种情况下的概率不同.
  - 原因: 第二种情况下, 我们多知道了一个条件.
- 第二种情况考虑:
  - 已知事件 B 发生的条件下, 事件 A 发生的概率, 记 $P(A|B)$.

2. 条件概率与事件的独立性

条件概率 - 例子

- 上述问题属于古典概型，样本空间中所含的样本点数量为 4.
- 假如已知事件 B 发生，即至少有一个女孩，那么可能的样本点只有 3 个，分别是：(男，女), (女，男), (女，女). 此时，满足"一个男孩，一个女孩"的事件数为 2，因此

$$P(A|B) = \frac{2}{3} = \frac{\frac{2}{4}}{\frac{3}{4}} = \frac{P(A \cap B)}{P(B)}.$$

- 虽然上面的公式以特例的形式引入，但是可以证明：对一般古典概型以及几何概率问题也成立.

定义 1.3.1

- 设 A, B 是两事件, 且 $P(B)>0$, 称

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

为事件 B 已经发生的条件下, 事件 A 发生的概率.

- 类似的, 定义事件 A 已经发生的条件下, 事件 B 发生的概率

$$P(B|A)=\frac{P(A\cap B)}{P(A)}\ (P(A)>0).$$

2. 条件概率与事件的独立性

条件概率 - 例子

- 袋中装有 16 个球, 其中 6 个是玻璃球, 另外 10 个是塑料球.
  - 玻璃球中有 2 个是红色, 4 个是蓝色.
  - 塑料球中 3 个是红色, 7 个是蓝色.
- 求取到蓝色球的条件下，发现取到的是玻璃球的概率.

2. 条件概率与事件的独立性

条件概率 - 例子

- 袋中装有 16 个球, 其中 6 个是玻璃球, 另外 10 个是塑料球.
  - 玻璃球中有 2 个是红色, 4 个是蓝色.
  - 塑料球中 3 个是红色, 7 个是蓝色.
- 求取到蓝色球的条件下，发现取到的是玻璃球的概率.
- $A = \{取到蓝色球\}$, 则 $P(A) = \frac{11}{16}$.
- $B = \{取到玻璃球\}$, 则 $P(B) = \frac{6}{16}$.
- $A \cap B = \{取到蓝色玻璃球\}$, $P(A \cap B) = \frac{4}{16}$
- $P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{4}{11}$.

条件概率 - 例子

- 在肝癌普查中发现，某地区的自然人群中，每十万人内平均有 40 人患原发性肝癌，有 34 人甲胎球蛋白高含量，有 32 人既患原发性肝癌又出现甲胎球蛋白高含量.

(1) 患原发性肝癌的人甲胎球蛋白呈现高含量的概率.

(2) 甲胎球蛋白高含量的人患原发性肝癌的概率.

条件概率 - 例子

- 在肝癌普查中发现，某地区的自然人群中，每十万人内平均有 40 人患原发性肝癌，有 34 人甲胎球蛋白高含量，有 32 人既患原发性肝癌又出现甲胎球蛋白高含量.

(1) 患原发性肝癌的人甲胎球蛋白呈现高含量的概率.

(2) 甲胎球蛋白高含量的人患原发性肝癌的概率.

- 记 $C=\{$患原发性肝癌$\}$, $D=\{$甲胎球蛋白高含量$\}$, 则有

$$P(C)=0.0004, P(D)=0.00034, P(C\cap D)=0.00032.$$

- 根据条件概率的定义有

$$P(D|C)=\frac{P(C\cap D)}{P(C)}=\frac{0.00032}{0.0004}=0.8$$

患原发性肝癌的人有 80% 其甲胎球蛋白呈现高含量.

- 根据条件概率的定义

$$P(C|D)=\frac{P(C\cap D)}{P(D)}=\frac{0.00032}{0.00034}=0.9412$$

甲胎球蛋白高含量的人有超过 94% 概率患原发性肝癌.

- 由于 D 的发生, C 发生的概率由 0.0004 上升到 0.9412.

条件概率 - 例子 - 教材 Page 20: 例 1.4.1

- 某单位 100 名员工做血压和肝功能指标检查. 结果有 95 人血压正常 (事件 A), 94 人肝功能正常 (事件 B), 92 人两项检查都正常. 在这个人群中, 如随机抽查一人, 求 $P(A)$, $P(B)$, $P(A|B)$, $P(B|A)$, $P(A\cap B)$.

2. 条件概率与事件的独立性

条件概率 - 例子 - 教材 Page 20: 例 1.4.1

- 某单位 100 名员工做血压和肝功能指标检查. 结果有 95 人血压正常 (事件 A), 94 人肝功能正常 (事件 B), 92 人两项检查都正常. 在这个人群中, 如随机抽查一人, 求 $P(A)$, $P(B)$, $P(A|B)$, $P(B|A)$, $P(A\cap B)$.

$$
\begin{aligned}
P(A) &= 0.95\\
P(B) &= 0.94\\
P(A\cap B) &= 0.92\\
P(A|B) &= \frac{P(A\cap B)}{P(B)}\\
P(B|A) &= \frac{P(A\cap B)}{P(A)}
\end{aligned}
$$

2. 条件概率与事件的独立性

观测

1. 事件发生的概率与条件有关，即与信息有关.
2. 条件概率 $P(B|A)$, $P(A|B)$ 仍是概率，只是在缩小的样本空间中分别计算事件 B 和事件 A 的概率，所以条件概率同样满足概率的公理化定义中的三条公理:
   2.1 非负性: $1 \geqslant P(A|B) \geqslant 0$.
   2.2 规范性: $P(\Omega|B) = 1$.
   2.3 可列可加性: $P\left(\bigcup\limits_{i=1}^{\infty} A_i \middle| B\right) = \sum\limits_{i=1}^{\infty} P(A_i|B)$, 其中 $A_1, A_2, \ldots, A_n, \ldots$ 为可列个两两互不相容事件.
3. 推论:

$$
\begin{aligned}
P(\emptyset|B) &= 0, \\
P(A|B) &= 1 - P(\overline{A}|B), \\
P(A_1 \cup A_2|B) &= P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B).
\end{aligned}
$$

概率的乘法公式/乘法定理

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

推广的乘法公式

$$\begin{aligned}P(A_1\cap A_2\cap\cdots\cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots\\ &\quad P(A_n|A_1\cap A_2\cap\cdots\cap A_{n-1})\\ &\quad (P(A_n|A_1\cap A_2\cap\cdots\cap A_{n-1})>0).\end{aligned}$$

乘法公式应用 - 例子 - 教材 Page 21: 例 1.4.2

- 在 100 件产品中有 3 件次品，现从中连取两次，每次取一件，取后不放回，求下列事件的概率:

  (1) 两次都是正品.
  (2) 一件正品，一件次品.

2. 条件概率与事件的独立性

乘法公式应用 - 例子 - 教材 Page 21: 例 1.4.2

- 在 100 件产品中有 3 件次品，现从中连取两次，每次取一件，取后不放回，求下列事件的概率：
  (1) 两次都是正品.
  (2) 一件正品，一件次品.
- 求解思路：

Step 1 通读题目，定义事件.

- $A_1 = \{$第 1 次取到正品$\}$, $A_2 = \{$第 2 次取到正品$\}$.

Step 2 确定所求的事件.

(1) $P(A_1 \cap A_2)$.

(2) $P\left((A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)\right)$.

Step 3 借助各种概率计算公式求解.

(1) $P(A_1 \cap A_2) = P(A_1) \cdot P(A_2|A_1) = \frac{97}{100} \cdot \frac{96}{99}$.

(2) $P\left((A_1 \cap \overline{A_2}) \cup (\overline{A_1} \cap A_2)\right) = P\left((A_1 \cap \overline{A_2}) + (\overline{A_1} \cap A_2)\right) =$
$P(A_1 \cap \overline{A_2}) + P(\overline{A_1} \cap A_2) =$
$P(A_1) \cdot P\left(\overline{A_2}|A_1\right) + P(\overline{A_1}) \cdot P\left(A_2|\overline{A_1}\right) = \frac{97}{100} \cdot \frac{3}{99} + \frac{3}{100} \cdot \frac{97}{99}$.

2. 条件概率与事件的独立性

推广的乘法公式应用 - 波利亚坛子模型

- 坛子中有 b 只黑球及 r 只红球, 随机取出一只, 把原球放回, 并加进与抽出球同色的球 c 只, 再摸第二次, 这样下去共摸了 n 次.
- **问题**: 前面 n_1 次出现黑球, 后面 $n_2 = n - n_1$ 次出现红球的概率?

2. 条件概率与事件的独立性

推广的乘法公式应用 - 波利亚坛子模型

- 坛子中有 b 只黑球及 r 只红球, 随机取出一只, 把原球放回, 并加进与抽出球同色的球 c 只, 再摸第二次, 这样下去共摸了 n 次.
- **问题**: 前面 n_1 次出现黑球, 后面 $n_2 = n - n_1$ 次出现红球的概率?

Step 1 通读题目, 定义事件.

- $A_1 = \{$第 1 次摸黑球$\}, \cdots, A_{n_1} = \{$第 n_1 次摸黑球$\}$.
- $A_{n_1+1} = \{$第 n_1+1 次摸红球$\}, \cdots, A_n = \{$第 n 次摸红球$\}$.

Step 2 确定所求的事件.

- $P(A_1 \cap A_2 \cap \cdots \cap A_n)$.

Step 3 借助各种概率计算公式求解.

$$\begin{aligned} P(A_1 \cap A_2 \cap \cdots \cap A_n) &= P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots \\ &\quad P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}). \end{aligned}$$

推广的乘法公式应用 - 波利亚坛子模型

- 坛子中有 b 只黑球及 r 只红球, 随机取出一只, 把原球放回, 并加进与抽出球同色的球 c 只, 再摸第二次, 这样下去共摸了 n 次.
- **问题**: 前面 n_1 次出现黑球, 后面 $n_2 = n - n_1$ 次出现红球的概率?
  - $A_1 = \{$第 1 次摸黑球$\}, \cdots, A_{n_1} = \{$第 n_1 次摸黑球$\}$.
  - $A_{n_1+1} = \{$第 n_1+1 次摸红球$\}, \cdots, A_n = \{$第 n 次摸红球$\}$.

Step 3 借助各种概率计算公式求解.

$$P(A_1) = \frac{b}{b+r},\ P(A_2|A_1) = \frac{b+c}{b+r+c},\ P(A_3|A_1\cap A_2) = \frac{b+2c}{b+r+2c}, \cdots,$$

$$P(A_{n_1}|A_1\cap\cdots\cap A_{n_1-1}) = \frac{b+(n_1-1)c}{b+r+(n_1-1)c},$$

$$P(A_{n_1+1}|A_1\cap\cdots\cap A_{n_1}) = \frac{r}{b+r+n_1c},$$

$$P(A_{n_1+2}|A_1\cap\cdots\cap A_{n_1+1}) = \frac{r+c}{b+r+(n_1+1)c}, \cdots,$$

$$P(A_n|A_1\cap\cdots\cap A_{n-1}) = \frac{r+(n_2-1)c}{b+r+(n-1)c}.$$

2. 条件概率与事件的独立性

推广的乘法公式应用 - 波利亚坛子模型

- 坛子中有 b 只黑球及 r 只红球，随即取出一只，把原球放回，并加进与抽出球同色的球 c 只，再摸第二次，这样下去共摸了 n 次.
- **问题**：前面 n_1 次出现黑球，后面 $n_2 = n - n_1$ 次出现红球的概率？
- 由推广的乘法公式

$$
\begin{aligned}
P(A_1 \cap A_2 \cap \cdots \cap A_n) = & \frac{b}{b+r} \cdot \frac{b+c}{b+r+c} \cdot \frac{b+2c}{b+r+2c} \\
& \cdots \frac{b+(n_1-1)c}{b+r+(n_1-1)c} \\
& \cdot \frac{r}{b+r+n_1 c} \cdot \frac{r+c}{b+r+(n_1+1)c} \\
& \cdots \frac{r+(n_2-1)c}{b+r+(n-1)c}.
\end{aligned}
$$

备注

- 注意答案仅与黑球及红球出现的次数有关，而与出现的顺序无关.

波利亚坛子模型的应用

- 波利亚坛子模型适用于群体增值和传染病的传播等现象.
- 模型的特征: 每一次取球后会增加与所取出的球具有相同颜色的球的个数, 而与所取出的球颜色不相同的球数目不增加. 因此, **每一次取出的球是什么颜色增加了下一次也取到这种颜色的球的概率**.
- 这是一个类似传染病传播的模型, 每发现一个患者, 就会增加再传染的概率.

独立事件 - 例子

- 一口袋中装有 a 只黑球和 b 只白球, 采用**有放回摸球**, 求:
  (1) 在已知第一次摸到黑球的条件下, 第二次摸出黑球的概率.
  (2) 第二次摸出黑球的概率.

SHANDONG UNIVERSITY 山东大学

独立事件 - 例子

- 一口袋中装有 a 只黑球和 b 只白球, 采用**有放回摸球**, 求:
  (1) 在已知第一次摸到黑球的条件下, 第二次摸出黑球的概率.
  (2) 第二次摸出黑球的概率.

Step 1 通读题目, 定义事件.

- $A=\{$第一次摸出黑球$\}$, $B=\{$第二次摸出黑球$\}$.

Step 2 确定所求的事件.

(1) $P(B|A)$.
(2) $P(B)$.

Step 3 借助各种概率计算公式求解.

$$P(A)=\frac{a}{a+b},\ P(A\cap B)=\frac{a^2}{(a+b)^2}\Rightarrow P(B|A)=\frac{P(A\cap B)}{P(A)}=\frac{a}{a+b}.$$

$$P(B)=P(A\cap B)+P(\overline{A}\cap B)=\frac{a^2}{(a+b)^2}+\frac{ba}{(a+b)^2}=\frac{a}{a+b}.$$

独立事件 - 例子

- 一口袋中装有 a 只黑球和 b 只白球, 采用**有放回摸球**, 求:
  (1) 在已知第一次摸到黑球的条件下, 第二次摸出黑球的概率.
  (2) 第二次摸出黑球的概率.

Step 3 借助各种概率计算公式求解.

$$P(A)=\frac{a}{a+b},\ P(AB)=\frac{a^2}{(a+b)^2}\Rightarrow P(B|A)=\frac{P(AB)}{P(A)}=\frac{a}{a+b}.$$

$$P(B)=P(AB)+P(\overline{A}B)=\frac{a^2}{(a+b)^2}+\frac{ba}{(a+b)^2}=\frac{a}{a+b}.$$

- $P(B|A)=P(B)\Rightarrow$ 事件A发生与否对事件B发生的概率没有影响.
- 直观原因: 采用有放回摸球.
- 事件 A 与事件 B 呈现出某种"独立性".

2. 条件概率与事件的独立性

定义 1.3.2

- 设事件 A 与 B, 若有

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

则称事件 A 与 B 相互独立.

- 等价定义: $P(A|B) = P(A)$ 或 $P(B|A) = P(B)$.

备注

- 若事件 A 和 B 相互独立, 则 A 关于 B 的条件概率等于无条件概率 $P(A)$, 即 B 的发生对 A 是否发生没有提供任何信息.

性质

- 必然事件 Ω 及不可能事件 $\emptyset$ 与任何事件独立.

2. 条件概率与事件的独立性

定理 1.3.3

- 若事件 A, B 相互独立, 则事件 $\overline{A}$ 与 B, A 与 $\overline{B}$, $\overline{A}$ 与 $\overline{B}$ 也相互独立.

证明

$$\begin{aligned}P(\overline{A}\cap B) &= P(B-A\cap B)=P(B)-P(A\cap B)\\ &= P(B)-P(A)P(B)=[1-P(A)]P(B)=P(\overline{A})P(B).\end{aligned}$$

独立事件 - 例子 - 抽签与顺序无关

- 一口袋中装有 a 只黑球和 b 只白球, 采用**不放回摸球**, 求:
  (1) 在已知第一次摸到黑球的条件下, 第二次摸出黑球的概率.
  (2) 第二次摸出黑球的概率.

独立事件 - 例子 - 抽签与顺序无关

- 一口袋中装有 a 只黑球和 b 只白球, 采用**不放回摸球**, 求:
  - (1) 在已知第一次摸到黑球的条件下, 第二次摸出黑球的概率.
  - (2) 第二次摸出黑球的概率.

Step 1 通读题目, 定义事件.

- $A = \{$第一次摸出黑球$\}$, $B = \{$第二次摸出黑球$\}$.

Step 2 确定所求的事件.

- $P(B|A)$ & $P(B)$.

Step 3 借助各种概率计算公式求解.

$$P(A) = \frac{a}{a+b}, P(A\cap B) = \frac{a}{a+b}\cdot\frac{a-1}{a+b-1} \Rightarrow P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{a-1}{a+b-1}.$$

$$P(\overline{A}\cap B) = \frac{b}{a+b}\cdot\frac{a}{a+b-1} \Rightarrow P(B) = P(A\cap B) + P(\overline{A}\cap B) = \frac{a}{a+b}.$$

- $P(B|A) \neq P(B) \Rightarrow$ **事件 B 与 A 不相互独立**.
- $P(A) = P(B) \Rightarrow$ **抽签与顺序无关**.

定义 1.3.3 (有限多个事件的独立性)

- 设 $A_1, A_2, \cdots, A_n$ 为 $n\ (n \geqslant 2)$ 个事件, 若对其中**任意** k $(2 \leqslant k \leqslant n)$ **个事件** $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ $(1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n)$, 满足等式

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k}),$$

则称事件 $A_1, A_2, \cdots, A_n$ 相互独立.

注意

- n 个事件**相互独立**.
  - 任意 2 个事件 A_{i_1}, A_{i_2}, $P(A_{i_1} \cap A_{i_2}) = P(A_{i_1}) \cdot P(A_{i_2})$.
  - 任意 3 个事件 $A_{i_1}, A_{i_2}, A_{i_3}$, $P(A_{i_1} \cap A_{i_2} \cap A_{i_3}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot P(A_{i_3})$.
  - $\cdots\ \cdots$
  - $P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1) \cdot P(A_2) \cdots P(A_n)$.
- n 个事件**相互独立**要求 $2^n - n - 1$ **个式子成立**.

2. 条件概率与事件的独立性

定义 1.3.3 (有限多个事件的独立性)

- 设 $A_1, A_2, \cdots, A_n$ 为 n $(n \geqslant 2)$ 个事件，若对其中**任意** k $(2 \leqslant k \leqslant n)$ **个事件** $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ $(1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n)$，满足等式

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdots P(A_{i_k}),$$

则称事件 $A_1, A_2, \cdots, A_n$ 相互独立.

注意

- n 个事件**相互独立**要求 $2^n - n - 1$ 个式子成立.
- n 个事件**两两相互独立**要求下面的 C_n^2 个式子成立.

$$P(A_i \cap A_j) = P(A_i) \cdot P(A_j) \quad (i \neq j; i, j = 1, 2, \cdots, n).$$

- n 个事件相互独立 $\Rightarrow$ n 个事件两两相互独立.
- n 个事件两两相互独立 $\nRightarrow$ n 个事件相互独立.

2. 条件概率与事件的独立性

注意

- n 个事件相互独立 $\Rightarrow$ n 个事件两两相互独立.
- n 个事件两两相互独立 $\nRightarrow$ n 个事件相互独立.

例子 - 三个事件的独立性

$$\text{相互独立}\left\{\begin{array}{l} P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C) \\ \left.\begin{array}{l} P(A\cap B)=P(A)\cdot P(B) \\ P(A\cap C)=P(A)\cdot P(C) \\ P(B\cap C)=P(B)\cdot P(C) \end{array}\right\}\text{两两相互独立} \end{array}\right.$$

例 1.3.6 - 两两独立但不相互独立 (伯恩斯坦反例)

- 设有四张卡片，其中三张分别涂有红色、白色和黑色，而剩下的一张同时涂有红、白、黑三色，从中任意抽取一张.
- $A=\{$抽到的卡片上有红色$\}$.
- $B=\{$抽到的卡片上有白色$\}$.
- $C=\{$抽到的卡片上有黑色$\}$.

$$P(A)=P(B)=P(C)=\frac{2}{4}=\frac{1}{2}$$

$$P(A\cap B)=P(A\cap C)=P(B\cap C)=\frac{1}{4}$$

- 因此 A, B, C**两两相互独立**.
- $P(A\cap B\cap C)=\frac{1}{4}\neq\frac{1}{8}=P(A)\cdot P(B)\cdot P(C)$.
- A, B, C**不相互独立**.

例 1.3.7 - 三个事件的独立性

- 设有八张卡片

1, 2, 3	1, 2	1, 2	3	3	3	2	1

- $A=\{$抽出的卡片上有 1$\}$.
- $B=\{$抽出的卡片上有 2$\}$.
- $C=\{$抽出的卡片上有 3$\}$.

$$P(A)=P(B)=P(C)=\frac{4}{8}=\frac{1}{2}, P(A\cap B)=\frac{3}{8}, P(A\cap C)=P(B\cap C)=\frac{1}{8}$$

$$P(A\cap B\cap C)=\frac{1}{8}=P(A)\cdot P(B)\cdot P(C)$$

$$P(A\cap B)=\frac{3}{8}\neq P(A)\cdot P(B), P(A\cap C)=\frac{1}{8}\neq P(A)\cdot P(C)$$

$$P(B\cap C)=\frac{1}{8}\neq P(B)\cdot P(C).$$

- $P(A\cap B\cap C)=P(A)\cdot P(B)\cdot P(C)$ **既不能**保证 A, B, C 独立，**也不能**保证 A, B, C 两两独立.

2. 条件概率与事件的独立性

无穷多个事件的独立性

- 称**无穷多个事件是相互独立的**, 如果其中任意有限多个事件都相互独立.

独立事件带来的好处

- 若事件是独立的, 那么许多概率计算问题可以简化.

相互独立事件至少发生其一的概率计算公式

- 若 $A_1, A_2, \cdots, A_n$ 是 n 个**相互独立**的事件, 由于

$$\overline{A_1 \cup A_2 \cup \cdots \cup A_n} = \overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n},$$

所以有

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - P(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}) \\ &= 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdots P(\overline{A_n}). \end{aligned}$$

独立性 + 逆事件 - 例子

- 加工某一种零件需要经过三道工序, 设三道工序的次品率分别为 2%, 1%, 5%, 并假设各道工序互不影响, 求加工出来的零件的次品率.

独立性 + 逆事件 - 例子

- 加工某一种零件需要经过三道工序，设三道工序的次品率分别为 2%，1%，5%，并假设各道工序互不影响，求加工出来的零件的次品率.

Step 1 通读题目，定义事件.

- $A_i = \{$第 i 道工序出现次品$\}$.

Step 2 确定所求的事件.

- $P(A_1 \cup A_2 \cup A_3)$.

Step 3 借助各种概率计算公式求解.

$$\begin{aligned} P(A_1 \cup A_2 \cup A_3) &= 1 - P\left(\overline{A_1 \cup A_2 \cup A_3}\right) = 1 - P\left(\overline{A_1} \cap \overline{A_2} \cap \overline{A_3}\right) \\ &= 1 - P\left(\overline{A_1}\right) \cdot P\left(\overline{A_2}\right) \cdot P\left(\overline{A_3}\right) \end{aligned}$$

独立性 + 逆事件 - 例子

- 设炮兵使用某型号的高射炮，每一门炮一发击中飞机的概率为 0.2，问需要多少门炮同时射击，才能以 90% 的把握一发就击中来犯的敌机？

独立性 + 逆事件 - 例子

- 设炮兵使用某型号的高射炮，每一门炮一发击中飞机的概率为 0.2，问需要多少门炮同时射击，才能以 90% 的把握一发就击中来犯的敌机？

Step 1 通读题目，定义事件.

- 需要 n 门炮，$A_i = \{$第 i 门炮击中敌机$\}$，$A = \{$敌机被击中$\}$.

Step 2 确定所求的事件.

- $P(A) = P(A_1 \cup A_2 \cup \cdots \cup A_n)$.

Step 3 借助各种概率计算公式求解.

$$\begin{aligned} P(A_1 \cup A_2 \cup \cdots \cup A_n) &= 1 - P(\overline{A_1} \cap \overline{A_2} \cap \cdots \cap \overline{A_n}) \\ &= 1 - P(\overline{A_1}) \cdot P(\overline{A_2}) \cdots P(\overline{A_n}). \\ 1 - 0.8^n &\geqslant 0.9. \end{aligned}$$

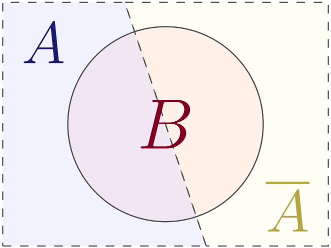
本节主要内容

1. 全概率公式.
2. 贝叶斯公式.

全概率公式 - 引入

- 从前面的例题中可以看到，有时为了计算 $P(B)$，可以找一个有关的事件 A，使用下面的公式

$$P(B) = P(A \cap B) + P(\overline{A} \cap B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$



全概率公式 - 引入

- 例：从装有 a 只黑球和 b 只白球的袋子中**不放回摸球**，求第二次摸得黑球的概率 $P(B)$.
  - $A = \{第一次摸得黑球\}$.

$$P(B) = \frac{a}{a+b} \cdot \frac{a-1}{a+b-1} + \frac{b}{a+b} \cdot \frac{a}{a+b-1} = \frac{a}{a+b}.$$

  - 摸球与顺序无关.
  - 先摸的人摸到黑球 $\Rightarrow$ 后摸的人可能处于“不利情况”，摸到黑球的概率降为 $\frac{a-1}{a+b-1}$.
  - 先摸的人摸到白球 $\Rightarrow$ 后摸的人可能处于“有利情况”，摸到黑球的概率提升为 $\frac{a}{a+b-1}$.
  - 最终结果是上述两种情况的加权平均，权重 $\frac{a}{a+b}$ 和 $\frac{b}{a+b}$ 可以看作处于“不利情况”和“有利情况”的概率.

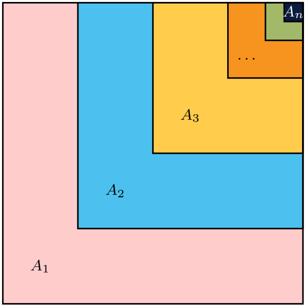
$$P(B) = P(A \cap B) + P(\overline{A} \cap B) = P(A) \cdot P(B|A) + P(\overline{A}) \cdot P(B|\overline{A}).$$

1. 全概率公式

完备事件组

- 设 n 个事件 $A_1, A_2, \ldots, A_n$, 若有
  1. $A_i\ (i=1,2,\ldots,n)$**两两互不相容**,
  2. $\sum\limits_{i=1}^{n} A_i = \Omega$,

  则称 $A_1, A_2, \ldots, A_n$ 构成一个**完备事件组**.

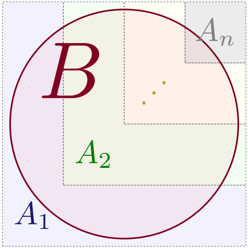


1. 全概率公式

定理 1.4.1 (全概率公式)

- 事件 $A_1, A_2, \ldots, A_n$ 构成一个**完备事件组**, 则对任一事件 B, 有

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i).$$



证明

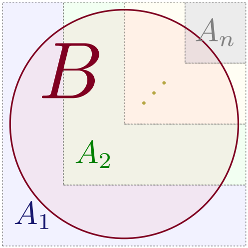
- $B = B \cap \Omega = B \cap \left(\sum_{i=1}^{n} A_i\right) = A_1 \cap B + A_2 \cap B + \cdots + A_n \cap B.$
- $P(B) = \sum_{i=1}^{n} P(A_i \cap B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i).$

1. 全概率公式

定理 1.4.1 (全概率公式)

- 事件 $A_1, A_2, \ldots, A_n$ 构成一个完备事件组，则对任一事件 B，有

$$P(B) = \sum_{i=1}^{n} P(A_i) \cdot P(B|A_i).$$



全概率公式 - 引申

- 可以将上述条件 $\sum_{i=1}^{n} A_i = \Omega$ **弱化**为 $\sum_{i=1}^{n} A_i \supset B$.
- 全概率公式反映了一种常见策略：分而治之/各个击破.

全概率公式 - 例子

- 设某厂有甲、乙、丙三个车间生产同一种产品，已知各车间的产量分别占全厂产量的 25%, 35%, 40%. 并且各车间的次品率依次为 5%, 4%, 2%. 现从该厂这批产品中任取一件，求这批产品的次品率.

全概率公式 - 例子

- 设某厂有甲、乙、丙三个车间生产同一种产品，已知各车间的产量分别占全厂产量的 25%, 35%, 40%. 并且各车间的次品率依次为 5%, 4%, 2%. 现从该厂这批产品中任取一件，求这批产品的次品率.

Step 1 通读题目，定义事件.

- $A_1 = \{$产品由甲车间生产$\}$, $A_2 = \{$产品由乙车间生产$\}$, $A_3 = \{$产品由丙车间生产$\}$.
- $B = \{$产品为次品$\}$.

Step 2 确定所求的事件.

- $P(B)$.

Step 3 借助各种概率计算公式求解.

- A_1, A_2, A_3 构成完备事件组，且 $P(A_1) = 25\%$, $P(A_2) = 35\%$, $P(A_3) = 40\%$.
- $P(B|A_1) = 5\%$, $P(B|A_2) = 4\%$, $P(B|A_3) = 2\%$.
- $P(B) = P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)$.

全概率公式 - 例子

- 甲袋中装有 2 个红球, 6 个白球.
- 乙袋中装有 5 个红球, 4 个白球.
- 现从甲袋中任取 3 个放入乙袋, 然后再从乙袋中任取一个球, 求取得红球的概率.

1. 全概率公式

全概率公式 - 例子

- 甲袋中装有 2 个红球, 6 个白球.
- 乙袋中装有 5 个红球, 4 个白球.
- 现从甲袋中任取 3 个放入乙袋, 然后再从乙袋中任取一个球, 求取得红球的概率.

Step 1 通读题目, 定义事件.

- $A_i = \{$甲袋中取出 3 个球有 i 个红球$\}$, $i = 0, 1, 2$.
- $B = \{$乙袋中取到红球$\}$.

Step 2 确定所求的事件.

- $P(B)$.

Step 3 借助各种概率计算公式求解.

- A_0, A_1, A_2 构成完备事件组, 且 $P(A_i) = \frac{C_2^i \cdot C_6^{3-i}}{C_8^3}$.
- $P(B|A_i) = \frac{5+i}{12}$.
- $P(B) = P(A_0) \cdot P(B|A_0) + P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2)$.

全概率公式 - 例子

- 甲、乙两人进行射击比赛，每回射击胜者得 1 分，并设在每回射击中甲胜的概率为 α，乙胜的概率为 β，$\alpha+\beta=1$，比赛进行到有一人比对手多 2 分为止，多得 2 分者最终获胜，求甲最终获胜的概率.

全概率公式 - 例子

- 甲、乙两人进行射击比赛，每回射击胜者得 1 分，并设在每回射击中甲胜的概率为 α，乙胜的概率为 β，$\alpha+\beta=1$，比赛进行到有一人比对手多 2 分为止，多得 2 分者最终获胜，求甲最终获胜的概率.

Step 1 通读题目，定义事件.

- $A_1=\{$前两轮甲均获胜$\}$，$A_2=\{$前两轮乙均获胜$\}$，$A_3=\{$前两轮甲、乙各胜一次$\}$.
- $B=\{$甲最终获胜$\}$.

Step 2 确定所求的事件.

- $P(B)$.

Step 3 借助各种概率计算公式求解.

- A_1, A_2, A_3 构成完备事件组，且 $P(A_1)=\alpha^2$，$P(A_2)=\beta^2$，$P(A_3)=2\alpha\beta$.
- $P(B|A_1)=1$，$P(B|A_2)=0$，$P(B|A_3)=P(B)$.
- $P(B)=P(A_1)\cdot P(B|A_1)+P(A_2)\cdot P(B|A_2)+P(A_3)\cdot P(B|A_3)$.

2. 贝叶斯公式

定理 1.4.2 (贝叶斯公式)

- 若事件 B**能且只能**与**两两互不相容**事件 $A_1, A_2, \ldots, A_n$ 之一同时发生，即 $B = \sum_{i=1}^{n} (B \cap A_i)$.
- 由条件概率的定义: $P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\cdot P(B|A_i)}{P(B)}$，有

$$P(A_i|B) = \frac{P(A_i)\cdot P(B|A_i)}{\sum_{i=1}^{n} P(A_i)\cdot P(B|A_i)}.$$

贝叶斯公式 - 注意

- 此处仅要求 $\sum_{i=1}^{n} A_i \supset B$，而不需要 $A_1, A_2, \ldots, A_n$ 构成完备事件组.

2. 贝叶斯公式

$$P(A_i|B) = \frac{P(A_i)\cdot P(B|A_i)}{\sum\limits_{i=1}^{n} P(A_i)\cdot P(B|A_i)}.$$

在实际应用中理解贝叶斯公式 - 人工智能辅助医疗诊断

- 诊断病人患疾病 $A_1, A_2, \ldots, A_n$ 中哪一种.
- 在对病人进行观察与检查后, 确定了某个指标 B(如: 体温、脉搏等), 希望利用这个指标来帮助诊断, 此时可以应用贝叶斯公式.
- 此时, 先验概率 $P(A_i)$ 指人患各种疾病的可能性大小, 以往的资料可以给出一些初步数据.
- 要确定 $P(B|A_i)$, 主要依靠医学知识.
- 利用贝叶斯公式可以计算出不同的 $P(A_i|B)$.
- 对应于较大 $P(A_i|B)$ 的病因 A_i 应多加考虑.

2. 贝叶斯公式

$$P(A_i|B) = \frac{P(A_i) \cdot P(B|A_i)}{\sum_{i=1}^{n} P(A_i) \cdot P(B|A_i)}.$$

理解贝叶斯公式

- 设 $A_1, A_2, \ldots, A_n$ 是导致试验结果的原因，称 $P(A_i)$ 为**先验概率**，反映了各种原因发生的可能性大小，一般是以往经验的总结，在**试验前已知**.
- 若试验中发生了事件 B，这个信息将有助于探讨事件发生的原因. 称条件概率 $P(A_i|B)$ 为**后验概率**，它反映了试验之后对各种原因发生的可能性大小的新知识.
- 贝叶斯公式借助先验概率计算后验概率.

贝叶斯公式 - 例子

- 设某厂有甲、乙、丙三个车间生产同一种产品，已知各车间的产量分别占全厂产量的 25%, 35%, 40%. 并且各车间的次品率依次为 5%, 4%, 2%. 若从该批产品中任取一件，该件是次品，问该次品由哪个车间生产的概率大？

2. 贝叶斯公式

贝叶斯公式 - 例子

- 设某厂有甲、乙、丙三个车间生产同一种产品，已知各车间的产量分别占全厂产量的 25%, 35%, 40%. 并且各车间的次品率依次为 5%, 4%, 2%. 若从该批产品中任取一件，该件是次品，问该次品由哪个车间生产的概率大?

Step 1 通读题目，定义事件.

- $A_1 = \{产品由甲车间生产\}$, $A_2 = \{产品由乙车间生产\}$, $A_3 = \{产品由丙车间生产\}$.
- $B = \{产品为次品\}$.

Step 2 确定所求的事件.

- $P(A_1|B)$, $P(A_2|B)$, $P(A_3|B)$.

Step 3 借助各种概率计算公式求解.

- A_1, A_2, A_3 构成完备事件组，且 $P(A_1) = 25\%$, $P(A_2) = 35\%$, $P(A_3) = 40\%$.
- $P(B|A_1) = 5\%$, $P(B|A_2) = 4\%$, $P(B|A_3) = 2\%$.
- $P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)} = \frac{P(A_1) \cdot P(B|A_1)}{P(A_1) \cdot P(B|A_1) + P(A_2) \cdot P(B|A_2) + P(A_3) \cdot P(B|A_3)}$.

贝叶斯公式 - 例子

- 甲胎蛋白免疫检测法 (AFP) 被普遍用于肝癌的早期诊断和普查. 已知肝癌患者经 AFP 诊断为肝癌的概率为 95%, 而未患肝癌通过 AFP 被误诊为肝癌的概率为 2%, 在人群中肝癌的发病率一般为 0.04%, 现有一人经 AFP 检验诊断为患肝癌, 求此人的确患肝癌的概率.

2. 贝叶斯公式

贝叶斯公式 - 例子

- 甲胎蛋白免疫检测法 (AFP) 被普遍用于肝癌的早期诊断和普查. 已知肝癌患者经 AFP 诊断为肝癌的概率为 95%, 而未患肝癌通过 AFP 被误诊为肝癌的概率为 2%, 在人群中肝癌的发病率一般为 0.04%, 现有一人经 AFP 检验诊断为患肝癌, 求此人的确患肝癌的概率.

Step 1 通读题目, 定义事件.

- $A=\{$患肝癌$\}$, $B=\{$经 AFP 检验诊断为患肝癌$\}$.

Step 2 确定所求的事件.

- $P(A|B)$.

Step 3 借助各种概率计算公式求解.

- 梳理已知条件: $P(B|A)=95\%$, $P(B|\overline{A})=2\%$, $P(A)=0.04\%$.
- $P(A|B)=\frac{P(A\cap B)}{P(B)}=\frac{P(A)\cdot P(B|A)}{P(A)\cdot P(B|A)+P(\overline{A})\cdot P(B|\overline{A})}$.

2. 贝叶斯公式

$$P(A_i|B) = \frac{P(A_i)\cdot P(B|A_i)}{\sum\limits_{i=1}^{n} P(A_i)\cdot P(B|A_i)}.$$

贝叶斯公式 - 例子

- 假定用血清甲胎球蛋白法诊断肝癌，$C = \{$被检验者患有肝癌$\}$，$A = \{$判断被检验者患有肝癌$\}$.
- $P(A|C) = 0.95$, $P(\overline{A}|\overline{C}) = 0.90$.
- 在自然人群中 $P(C) = 0.0004$.
- 问题：现在若有一人被此检验法诊断为患有肝癌，求此人真正患有肝癌的概率 $P(C|A)$?

2. 贝叶斯公式

$$P(A_i|B)=\frac{P(A_i)\cdot P(B|A_i)}{\sum\limits_{i=1}^{n}P(A_i)\cdot P(B|A_i)}.$$

贝叶斯公式 - 例子

- 假定用血清甲胎球蛋白法诊断肝癌，$C=\{$被检验者患有肝癌$\}$，$A=\{$判断被检验者患有肝癌$\}$.
- $P(A|C)=0.95$, $P(\overline{A}|\overline{C})=0.90$.
- 在自然人群中 $P(C)=0.0004$.
- 问题：现在若有一人被此检验法诊断为患有肝癌，求此人真正患有肝癌的概率 $P(C|A)$?
- 由贝叶斯公式

$$\begin{aligned}P(C|A)&=\frac{P(C)\cdot P(A|C)}{P(C)\cdot P(A|C)+P(\overline{C})\cdot P(A|\overline{C})}\\&=\frac{0.0004\times 0.95}{0.0004\times 0.95+0.9996\times 0.10}=0.0038.\end{aligned}$$

贝叶斯公式 - 例子

- 假定用血清甲胎球蛋白法诊断肝癌，$C = \{$被检验者患有肝癌$\}$，$A = \{$判断被检验者患有肝癌$\}$.
- $P(A|C) = 0.95$, $P(\overline{A}|\overline{C}) = 0.90$. 在自然人群中 $P(C) = 0.0004$.
- 为什么 $P(C|A)$ 只有 0.0038?

贝叶斯公式 - 例子

- 假定用血清甲胎球蛋白法诊断肝癌, $C=\{$被检验者患有肝癌$\}$, $A=\{$判断被检验者患有肝癌$\}$.
- $P(A|C)=0.95$, $P(\overline{A}|\overline{C})=0.90$. 在自然人群中 $P(C)=0.0004$.
- 为什么 $P(C|A)$ 只有 0.0038?
- 原因: 先验概率 $P(C)$ 太小了.
- 每十万人中有 40 人患肝癌, $P(A|C)=0.95 \Rightarrow 38$ 个人判断正确, 漏掉 2 人.
- 99960 个正常人, $P(\overline{A}|\overline{C})=0.90$, 误判 9996 人.
- 在 10034(=9996+38) 个潜在患者中, 只有 $\frac{38}{10034}=0.0038$ 的可能性真的患有肝癌.

2. 贝叶斯公式

贝叶斯公式 - 例子

- 如果把调查人群换为肝癌的可疑人群, 如: 甲胎球蛋白高含量者.
- 在这一群体中, 患肝癌的概率为 $P(C) = 0.9412$.
- 由贝叶斯公式

$$P(C|A) = \frac{0.9412 \times 0.95}{0.9412 \times 0.95 + 0.0588 \times 0.1} = 0.9935.$$

可以看到后验概率提升了.

注意

- 后验概率的大小受先验概率选取的影响.

§1.5 伯努利概型

伯努利试验

- 对试验感兴趣的是试验中某事件 A 是否发生.
  - 在产品抽样检查中注意的是抽到废品, 还是抽到正常产品.
  - 在掷硬币时关心出现正面还是出现反面.
  - 股票市场中关心涨还是跌.
- 这种 **只有两个可能结果**的试验称为**伯努利试验**.

注意

- 有些试验的结果不止两个, 但也可看作伯努利试验.
  - 在电报传输中, 既要传送字母 A, B, . . ., Z 等, 又要传送其他符号. 但是假如我们所关心的只是 **字母在传送中所占的百分比**, 而不再区别到底是哪一个字母, 则可以把**出现字母当作是成功**, 出现其他符号一律当作是失败.
  - 显像管的寿命可以是不小于 0 的任一数值, 但是有时根据需要, 我们可以把**寿命大于 50000 小时的显像管当作合格品**, 其余都作为次品.
- **伯努利试验 $\Leftrightarrow$ 样本空间中仅含两个样本点.**

§1.5 伯努利概型

伯努利试验

- 在伯努利试验中，需要给出概率
$$P(A)=p,\ P(\overline{A})=q.$$
- 显然 $p\geqslant 0,\ q\geqslant 0$，且 $p+q=1$.

n 重伯努利试验

- 现在考虑n **次独立重复的伯努利试验**，这里的“重复”指在每次试验中事件 A 与事件 $\overline{A}$ 出现的概率都保持不变. 这试验称为 n **重伯努利试验**，记作 E^n.

n 重伯努利试验的四个特征

1. 每次试验至多出现**两个可能结果之一**：A 或 $\overline{A}$.
2. A 在每次试验中出现的**概率保持不变**.
3. 各次试验相互**独立**.
4. 共进行n **次**试验

§1.5 伯努利概型

n 重伯努利试验

- 现在考虑n **次独立重复的伯努利试验**, 这里的“重复”指在每次试验中事件 A 与事件 $\overline{A}$ 出现的概率都保持不变. 这试验称为 n **重伯努利试验**, 记作 E^n.

n 重伯努利试验的样本点

- n 重伯努利试验 E^n 样本点: $(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n)$.
- $\hat{A}_i$ 可取 A_i 或 $\overline{A_i}$, 分别表示第 i 次试验中出现 A 或 $\overline{A}$.
- 将样本点简记为 $\hat{A}_1\hat{A}_2\cdots\hat{A}_n$.
  - 例: $A_1A_2\cdots A_{n-1}\overline{A_n}$ 表示前 $n-1$ 次试验均出现事件 A, 第 n 次出现事件 $\overline{A}$.
- **样本点共有 2^n 个**, 样本空间是有限集.
  - **不一定等可能**, 注意与古典概型区分.
- 样本点的任意子集可以作为事件.

n 重伯努利试验样本点的概率

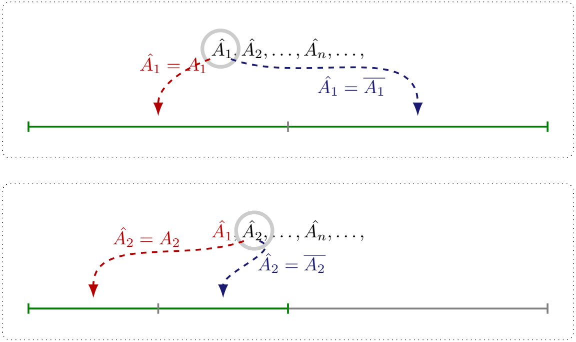
- 样本点的概率, 主要看其中 A 或 $\overline{A}$ 出现的次数. 若其中有 ℓ 个 A, $n-\ell$ 个 $\overline{A}$, 则有

$$P(\hat{A}_1\hat{A}_2\cdots\hat{A}_n) = p^{\ell}\cdot q^{n-\ell}.$$

- 特别的, $P(A_1A_2\cdots\overline{A_n}) = p^{n-1}\cdot q$.
- 一般事件的概率由它所含样本点的概率求和得到.

可列重伯努利试验 E^∞

- 样本点: $(\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_n, \ldots,)$, 其中 $\hat{A}_i$ 可取 A_i 或 $\overline{A_i}$.
- 样本空间不再有限, 甚至不可列/不可数.
- 样本空间可与 $[0,1]$ 区间进行一一对应.



§1.5 伯努利概型

伯努利试验 - 重要性

- “在同样的条件下进行重复试验”的一种数学模型.

理论意义 **最早**研究的模型之一, 也是得到**最多**研究的模型之一.

实际意义 广泛的实际应用, 例如在工业产品质量检查中.

- 伯努利概型中常见事件的概率.
  1. 伯努利分布
  2. 二项分布
  3. 几何分布
  4. 帕斯卡分布

§1.5 伯努利概型

伯努利分布

- **只进行一次**伯努利试验，则或是事件 A 出现，或是事件 $\overline{A}$ 出现，其概率为

$$P(A) = p, \ \ P(\overline{A}) = q.$$

称为**伯努利分布**，这是最简单的情况.

二项分布

- n重伯努利试验中事件 A 出现k次的概率，记为 $B(k;n,p)$.
- $B_k=\{n$ 重伯努利试验中事件 A 正好出现 k 次$\}$,

$$B_k=A_1A_2\cdots A_k\overline{A_{k+1}}\cdots\overline{A_n}+\cdots+\overline{A_1A_2}\cdots\overline{A_{n-k}}A_{n-k+1}\cdots A_n.$$

- B_k 包含 C_n^k 种可能的情况，每种情况的概率为 $p^k\cdot q^{n-k}$.
- 利用概率的可加性，

$$P(B_k)=C_n^k\cdot p^k\cdot q^{n-k},$$
$$\Rightarrow B(k;n,p)=C_n^k\cdot p^k\cdot q^{n-k},k=0,1,2,\ldots,n.$$

§1.5 伯努利概型

二项分布 - 性质

- 注意到 $B(k;n,p)$ 是二项式 $(q+p\cdot s)^n$ 展开式中 s^k 项的次数，这也是‘二项分布’这一名称的由来.
- 特别的
$$\sum_{k=0}^{n} B(k;n,p) = \sum_{k=0}^{n} C_n^k \cdot p^k \cdot q^{n-k} = (q+p)^n = 1.$$

二项分布 - 例子

- 若在 N 件产品中有 M 件废品，现进行 n 次有放回的抽样检查，共抽得 k 件废品的概率？

二项分布 - 例子

- 若在 N 件产品中有 M 件废品，现进行 n 次有放回的抽样检查，共抽得 k 件废品的概率？

Step 1 通读题目，定义事件.

- $A = \{$一次试验中抽到废品$\}$, $P(A) = \frac{M}{N}$.

Step 2 确定所求的事件.

- $B\left(k; n, \frac{M}{N}\right)$.

Step 3 借助各种概率计算公式求解.

$$B\left(k; n, \frac{M}{N}\right) = C_n^k \cdot \left(\frac{M}{N}\right)^k \cdot \left(1 - \frac{M}{N}\right)^{n-k}.$$

二项分布 - 例子

- 对某种药物的疗效进行研究，设这种药物对某种疾病的有效率为 $p = 0.8$，现有 10 名患此种疾病的患者同时服用此药，求其中至少有 6 名患者服药有效的概率.

§1.5 伯努利概型

二项分布 - 例子

- 对某种药物的疗效进行研究, 设这种药物对某种疾病的有效率为 $p = 0.8$, 现有 10 名患此种疾病的患者同时服用此药, 求其中至少有 6 名患者服药有效的概率.

Step 1 通读题目, 定义事件.

- $A = \{$一名患者服药有效$\}$, $P(A) = 0.8$.

Step 2 确定所求的事件.

- $B(6; 10, 0.8) + B(7; 10, 0.8) + B(8; 10, 0.8) + B(9; 10, 0.8) + B(10; 10, 0.8)$.

Step 3 借助各种概率计算公式求解.

$$B(k; 10, 0.8) = C_{10}^{k} \cdot (0.8)^{k} \cdot (1 - 0.8)^{10-k}.$$

二项分布 - 例子

- **一大批**某型号的电子器件，已知其一级品率为 0.3，现从中随机抽查 20 只，问其中有一级品的概率是多少？

二项分布 - 例子

- **一大批**某型号的电子器件，已知其一级品率为 0.3，现从中随机抽查 20 只，问其中有一级品的概率是多少？

Step 1 通读题目，定义事件.

- $A = \{\text{一只为一级品}\}$，$P(A) = 0.3$.

Step 2 确定所求的事件.

- $\sum_{k=1}^{20} B(k; 20, 0.3)$.

Step 3 借助各种概率计算公式求解.

$$\begin{aligned}\sum_{k=0}^{20} B(k;20,0.3) &= 1. \\ \sum_{k=1}^{20} B(k;20,0.3) &= 1 - B(0;20,0.3) = 1 - 0.7^{20}.\end{aligned}$$

§1.5 伯努利概型

几何分布

- 伯努利试验中**首次成功**出现在第k次试验的概率.
- 必须而且只需在前 $k-1$ 次试验中都出现 $\overline{A}$, 而第 k 次试验出现 A, 记

$$W_k = \overline{A_1 A_2} \cdots \overline{A_{k-1}} A_k.$$

- 根据试验的独立性,

$$P(W_k) = P(\overline{A_1}) \cdot P(\overline{A_2}) \cdots P(\overline{A_{k-1}}) \cdot P(A_k) = q^{k-1} \cdot p.$$

$$G(k; p) = q^{k-1} \cdot p, k = 1, 2, \ldots.$$

- $G(k; p)$ 是几何级数的一般项, 因此这一分布称为**几何分布**.

§1.5 伯努利概型

几何分布 - 性质

- 根据几何级数

$$\sum_{k=1}^{\infty} G(k;p) = \sum_{k=1}^{\infty} q^{k-1} \cdot p = p \cdot \frac{1}{1-q} = 1.$$

- 几何分布给出了等待事件 A 出现共试验了 k 次的概率.

几何分布 - 例子

- 一个人要开门，他共有 n 把钥匙，其中仅有一把是能开这门的. 他随机地选取一把钥匙开门，即在每次试开时每一把钥匙都以概率 $\frac{1}{n}$ 被使用，这人在第 s 次试开时才首次成功的概率是多少?

SHANDONG UNIVERSITY 山东大学

几何分布 - 性质

- 根据几何级数

$$\sum_{k=1}^{\infty} G(k;p) = \sum_{k=1}^{\infty} q^{k-1} \cdot p = p \cdot \frac{1}{1-q} = 1.$$

- 几何分布给出了等待事件 A 出现共试验了 k 次的概率.

几何分布 - 例子

- 一个人要开门, 他共有 n 把钥匙, 其中仅有一把是能开这门的. 他随机地选取一把钥匙开门, 即在每次试开时每一把钥匙都以概率 $\frac{1}{n}$ 被使用, 这人在第 s 次试开时才首次成功的概率是多少?

$$G\left(s;\frac{1}{n}\right) = \left(\frac{n-1}{n}\right)^{s-1} \cdot \frac{1}{n}.$$

帕斯卡分布

- 考虑要多长时间才会出现**第 r 次成功**.
- 若**第 r 次成功**发生在第k次, 那么必有 $k \geqslant r$.
- $C_k = \{$第 r 次成功发生在第 k 次试验$\}$, $f(k;r,p) \triangleq P(C_k)$.
- C_k 发生当且仅当前 $k-1$ 次试验中有 $r-1$ 次成功, $k-r$ 次失败, 而第 k 次试验成功. 根据事件的独立性

$$P(C_k) = C_{k-1}^{r-1} \cdot p^{r-1} \cdot q^{k-r} \cdot p = C_{k-1}^{r-1} \cdot p^{r} \cdot q^{k-r},$$

$$f(k;r,p) = C_{k-1}^{r-1} \cdot p^{r-1} \cdot q^{k-r} \cdot p = C_{k-1}^{r-1} \cdot p^{r} \cdot q^{k-r}, k = r, r+1, \ldots.$$

牛顿广义二项式定理

$$(x+y)^\alpha = \sum_{k=0}^{\infty} C_\alpha^k \cdot x^{\alpha-k} \cdot y^k, \forall \alpha \in \mathbb{R},$$

$$C_\alpha^k = \frac{\alpha \cdot (\alpha-1) \cdots (\alpha-k+1)}{k!}.$$

帕斯卡分布 - 性质

$$
\begin{aligned}
\sum_{k=r}^{\infty} f(k;r,p) &= \sum_{k=r}^{\infty} C_{k-1}^{r-1}\cdot p^r\cdot q^{k-r} = \sum_{\ell=0}^{\infty} C_{\ell+r-1}^{r-1}\cdot p^r\cdot q^{\ell} (k \triangleq \ell + r)\\
&= \sum_{\ell=0}^{\infty} C_{\ell+r-1}^{\ell}\cdot p^r\cdot q^{\ell} = \sum_{\ell=0}^{\infty} C_{-r}^{\ell}\cdot(-1)^{\ell}\cdot p^r\cdot q^{\ell}\\
C_{\ell+r-1}^{\ell} &= \frac{(\ell+r-1)\cdot(\ell+r-1)\cdots(\ell+r-1-(\ell-1))}{\ell!}\\
&= \frac{r\cdot(r+1)\cdots(r+\ell-1)}{\ell!}\\
&= (-1)^{\ell}\frac{(-r)\cdot(-r-1)\cdots(-r-\ell+1)}{\ell!}\\
&= p^r\cdot\sum_{\ell=0}^{\infty} C_{-r}^{\ell}\cdot(-q)^{\ell} = p^r\cdot(1-q)^{-r} = 1.
\end{aligned}
$$

- $f(k;r,p)$ 称为帕斯卡分布.
- 特别的, 当 $r=1$ 时, 帕斯卡分布变成几何分布.

帕斯卡分布 - 例子 - 分赌注问题

- 甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙 s 局 $(r < t, s < t)$ 时，因故不得不终止，问如何分配这些赌注才公平合理？

帕斯卡分布 - 例子 - 分赌注问题

- 甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙 s 局 $(r < t, s < t)$ 时，因故不得不终止，问如何分配这些赌注才公平合理？
- $A = \{\text{一局中甲获胜}\}$, $P(A) = p$.
- 记 $n = t - r$, $m = t - s$ 为甲和乙为达到最后胜利所需再胜的局数.
- 记甲最终获胜的概率为 $p_{\text{甲}}$.
- 赌注以 $p_{\text{甲}} : (1 - p_{\text{甲}})$ 分配是公平合理的.
- $p_{\text{甲}}$ 等于在**出现 m 次 $\overline{A}$ 之前出现 n 次 A**的概率.

帕斯卡分布 - 例子 - 分赌注问题

- 甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙 s 局 $(r < t, s < t)$ 时，因故不得不终止，问如何分配这些赌注才公平合理？
- $A = \{$一局中甲获胜$\}$, $P(A) = p$.
- 记 $n = t - r$, $m = t - s$ 为甲和乙为达到最后胜利所需再胜的局数.
- $p_{甲}$ 等于在**出现 m 次 $\overline{A}$ 之前出现 n 次 A**的概率.
- **解法 1:** 利用帕斯卡分布，分别计算
  - 甲第 n 次成功出现在第 n 次试验.
  - 甲第 n 次成功出现在第 $n+1$ 次试验.
  - $\cdots\cdots$.
  - 甲第 n 次成功出现在第 $n+m-1$ 次试验.

$$p_{甲} = \sum_{k=0}^{m-1} C_{n+k-1}^{n-1} \cdot p^n \cdot q^k.$$

帕斯卡分布 - 例子 - 分赌注问题

- 甲、乙两个赌徒按某种方式下注赌博，说定先胜 t 局者将赢得全部赌注，但进行到甲胜 r 局，乙 s 局 $(r < t, s < t)$ 时，因故不得不终止，问如何分配这些赌注才公平合理？
- $A = \{\text{一局中甲获胜}\}$, $P(A) = p$.
- 记 $n = t - r$, $m = t - s$ 为甲和乙为达到最后胜利所需再胜的局数.
- **解法 2:** 可以证明，再赌 $n + m - 1$ 局一定可以决定胜负，所以甲为了取得胜利，只需在后续的 $n + m - 1$ 局中至少胜 n 局. 由二项分布

$$p_{\text{甲}} = \sum_{k=n}^{n+m-1} C_{n+m-1}^{k} \cdot p^k \cdot q^{n+m-1-k}.$$

帕斯卡分布 - 例子 - 巴拿赫火柴盒问题

- 一个人的左、右衣袋中各放有一盒装有 N 根火柴的火柴盒，每次抽烟时任取一盒用一根，求发现一盒用光时，另一盒有 r 根的概率.

帕斯卡分布 - 例子 - 巴拿赫火柴盒问题

- 一个人的左、右衣袋中各放有一盒装有 N 根火柴的火柴盒，每次抽烟时任取一盒用一根，求发现一盒用光时，另一盒有 r 根的概率.
- 首先考虑左边先空右边剩 r 根.
  - 左边摸过 $N+1$ 次.
    - 前 N 次用去 N 根火柴.
    - 最后一次发觉火柴盒是空的.
  - 右边摸过 $N-r$ 次.
- 求：第 $N+1$ 次摸左边发生在第 $2N-r+1$ 次抽烟时.

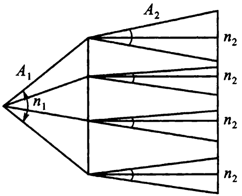
$$f(2N-r+1;N+1,\frac{1}{2}) = C_{2N-r}^{N}\cdot\left(\frac{1}{2}\right)^{N+1}\cdot\left(\frac{1}{2}\right)^{N-r}$$

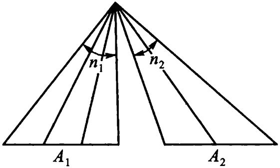
- 因此，最终的概率为 (右侧先空同理)

$$2\cdot f(2N-r+1;N+1,\frac{1}{2}) = C_{2N-r}^{N}\cdot\left(\frac{1}{2}\right)^{2N-r}.$$

组合分析公式推导原理

- **乘法原理：** 若进行 A_1 过程有 n_1 种方法，进行 A_2 过程有 n_2 种方法，则进行 A_1 过程后再接着进行 A_2 过程共有 $n_1 \times n_2$ 种方法.
- **加法原理：** 若进行 A_1 过程有 n_1 种方法，进行 A_2 过程有 n_2 种方法，假定 A_1 过程与 A_2 过程是并行的，则进行过程 A_1 或过程 A_2 的方法共有 $n_1 + n_2$ 种 (注意：不重不漏).
- 这二条原理可以推广到多个过程的场合.





排列

- 从包含有 n 个**不同**元素的总体中取出 r 个来进行**排列**, 这时既要考虑到**取出的元素**也要顾及其**取出顺序**.
- 这种排列可分为两类:
  1. **有放回地选取**: 这时每次选取是在全体元素中进行, 同一元素可被重复选中.

     $\Rightarrow$ **有重复的排列**, 总数共 n^r 种.
  2. **不放回选取**这时一个元素一旦被取出便立刻从总体中除去, 因此每个元素至多被选中一次, 必有 $r \leqslant n$.

     $\Rightarrow$ **选排列**, 总数共 $n \cdot (n-1) \cdots (n-r+1) = A_n^r$ 种.
     - 特别的, n 个不同元素的**全排列**数为 $n!$.

组合

1. 从 n 个**不同**的元素中取出 r 个元素而**不考虑其顺序**, 称为**组合**, 总数为

$$C_n^r = \frac{n!}{(n-r)! \cdot r!}.$$

这里 C_n^r 称为**二项式系数**, 是下列二项展开式的系数:

$$(a+b)^n = \sum_{r=0}^{n} C_n^r \cdot a^r \cdot b^{n-r}.$$

组合

2. 若 $r_1 + r_2 + \cdots + r_k = n$, 把 n **个不同的元素分成 k 个部分**, 第一部分 r_1 个, 第二部分 r_2 个, ..., 第 k 个部分 r_k 个, 不同的分法有

$$\frac{n!}{r_1! \cdot r_2! \cdots r_k!} = C_n^{r_1} \cdot C_{n-r_1}^{r_2} \cdots C_{r_{k-1}+r_k}^{r_{k-1}}.$$

上式中的数称为**多项系数**, 因为它是 $(x_1 + x_2 + \cdots + x_k)^n$ 展开式中 $x_1^{r_1} x_2^{r_2} \cdots x_k^{r_k}$ 的系数.

- 特别的, 当 $k = 2$ 时, 即为二项系数.

补充内容: 概率论常用的组合分析公式

组合

3. 若 n 个元素中: n_1 个带足标 "1", n_2 个带足标 "2", …, n_k 个带足标 "k", 且 $n_1+n_2+\cdots+n_k=n$.
   - ▷ 从这 n 个元素中取出 r 个, 使得带有足标 "i" 的元素有 r_i 个 $(1\leqslant i\leqslant k)$, 而 $r_1+r_2+\cdots+r_k=r$, 这时不同总数为

$$C_{n_1}^{r_1}\cdot C_{n_2}^{r_2}\cdots C_{n_k}^{r_k}.$$

补充内容：概率论常用的组合分析公式

组合

4. 从 n 个不同的元素中有重复地取 r 个，不计顺序，则不同的取法有 C_{n+r-1}^{r} 种，这个数称为**有重复组合数**.

- ▷ n 个不同的元素 $\Rightarrow$ n 个不同的盒子 $\square$.
- ▷ 取 r 个元素 $\Rightarrow$ r 个小球 $\bigcirc$.

$$\square\cdots \qquad \underbrace{\square\bigcirc}_{\text{表示将小球放在前面的盒子里}} \qquad \cdots$$

- ▷ 由于小球一定要放到盒子里 $\Rightarrow$ 第一个元素一定是盒子.
- ▷ 剩下的元素从中取 r 个位置.

补充内容: 概率论常用的组合分析公式

关于二项系数的一些公式

1. 在二项系数的定义中, 约定 $0! = 1$.
2. 对一切 $0 \leqslant k \leqslant n$, 有 $C_n^k = C_n^{n-k}$.
3. 对正整数 n 及 k, 若 $k > n$, 则 $C_n^k = 0$.
4. 二项式展开

$$(1+x)^n = \sum_{r=0}^{n} C_n^r \cdot x^r.$$

令 $x = 1$, 即有

$$C_n^0 + C_n^1 + \cdots + C_n^n = 2^n.$$

补充内容：概率论常用的组合分析公式

关于二项系数的一些公式

4. 对一切正整数 a, b, 成立

$$C_a^0 \cdot C_b^n + C_a^1 \cdot C_b^{n-1} + \cdots + C_a^n \cdot C_b^0 = C_{a+b}^n.$$

特别的

$$C_n^0 \cdot C_n^n + C_n^1 \cdot C_n^{n-1} + \cdots + C_n^n \cdot C_n^0 = C_{2n}^n,$$
$$\Rightarrow (C_n^0)^2 + (C_n^1)^2 + \cdots + (C_n^n)^2 = C_{2n}^n.$$

关于二项系数的一些公式

5. 把排列公式 A_n^r 推广到 r 是正整数而 n **是任意实数** x 的场合，有

$$A_x^r = x \cdot (x-1) \cdot (x-2) \cdots (x-r+1),$$

类似的，定义

$$C_x^r = \frac{A_x^r}{r!} = \frac{x \cdot (x-1) \cdot (x-2) \cdots (x-r+1)}{r!}.$$

规定 $C_x^0 = 1$.

- 可以证明

$$C_{-a}^k = (-1)^k \cdot C_{a+k-1}^k.$$

- 对任意实数 α，有牛顿二项式

$$(1+x)^\alpha = \sum_{r=0}^{\infty} C_\alpha^r \cdot x^r.$$

补充练习 1

- 在幸运 37 选 7 福利彩票中，每期从 1, 2, $\cdots$, 37 中开出7 个基本号码和一个特殊号码，彩民们在购买每一张彩票时都预先选定 7 个号码．规定 7 个基本号码全部选中者获一等奖，选中 6 个基本号码及特殊号码者获二等奖．试求购买一张彩票中一等奖的概率 p_1 及中二等奖的概率 p_2．

补充练习 1

- 在幸运 37 选 7 福利彩票中，每期从 1, 2, ⋯, 37 中开出7 个基本号码和一个特殊号码，彩民们在购买每一张彩票时都预先选定 7 个号码. 规定 7 个基本号码全部选中者获一等奖，选中 6 个基本号码及特殊号码者获二等奖. 试求购买一张彩票中一等奖的概率 p_1 及中二等奖的概率 p_2.
- 解:

$$
\begin{aligned}
p_1 &= \frac{C_7^7}{C_{37}^7} = 9.713 \times 10^{-8}. \\
p_2 &= \frac{C_7^6 \cdot C_1^1}{C_{37}^7} = 6.8 \times 10^{-7}.
\end{aligned}
$$

补充练习 2

- 甲有 $n+1$ 个硬币，乙有 n 个硬币，双方投掷之后进行比较，求甲掷出的正面数比乙掷出的正面数多的概率.

补充练习

补充练习 2

- 甲有 $n+1$ 个硬币, 乙有 n 个硬币, 双方投掷之后进行比较, 求甲掷出的正面数比乙掷出的正面数多的概率.
- 解: 这个问题初看非经繁复计算难求答案, 但是若充分利用它特有的对称并选择适当的样本空间, 则能迅速求解.
- 记 $A=\{$甲的正面数 $>$ 乙的正面数$\}$, 则
  $\overline{A}=\{$甲的正面数 $\leqslant$ 乙的正面数$\}$.
- 考虑到甲只比乙多一个硬币
  $\overline{A}=\{$甲的反面数 $>$ 乙的反面数$\}$.
- 由硬币的对称性, $P(A)=P(\overline{A})=\frac{1}{2}$.

补充练习 3

- 一袋中装有 10 个号码球, 分别标有 $1 \sim 10$ 号，现从袋中任取 3 个球，记录下其号码，求:
  (1) 最小号码为 5 的概率.
  (2) 最大号码为 5 的概率.
  (3) 中间的号码恰为 5 的概率.

补充练习

补充练习 3

- 一袋中装有 10 个号码球, 分别标有 1 ~ 10 号，现从袋中任取 3 个球，记录下其号码，求:
  (1) 最小号码为 5 的概率.
  (2) 最大号码为 5 的概率.
  (3) 中间的号码恰为 5 的概率.
- $P_1 = \frac{C_5^2}{C_{10}^3} = \frac{1}{12}$.
- $P_2 = \frac{C_4^2}{C_{10}^3} = \frac{1}{20}$.
- $P_3 = \frac{C_4^1 \cdot C_5^1}{C_{10}^3} = \frac{1}{6}$.

补充练习 4

- 从 6 双不同的手套中任取 4 只，问其中有一双配对的概率？

补充练习 4

- 从 6 双不同的手套中任取 4 只，问其中有一双配对的概率?

$$P=\frac{6\times C_4^2\times 2\times 1\times 10\times 8}{A_{12}^4}=\frac{16}{33}.$$

补充练习

补充练习 4

- 从 6 双不同的手套中任取 4 只，问其中有一双配对的概率？

$$P = \frac{6 \times C_4^2 \times 2 \times 1 \times 10 \times 8}{A_{12}^4} = \frac{16}{33}.$$

补充练习 5

- m 个男孩和 n 个女孩 $(n \leqslant m)$ 随机地沿着圆桌坐下，试求任意两个女孩都不相邻的概率？

补充练习

补充练习 4

- 从 6 双不同的手套中任取 4 只，问其中有一双配对的概率?

$$P = \frac{6 \times C_4^2 \times 2 \times 1 \times 10 \times 8}{A_{12}^4} = \frac{16}{33}.$$

补充练习 5

- m 个男孩和 n 个女孩 $(n \leqslant m)$ 随机地沿着圆桌坐下，试求任意两个女孩都不相邻的概率?

$$P = \frac{\frac{m!}{m} \cdot C_m^n \times n!}{\frac{(m+n)!}{m+n}} = \frac{C_m^n}{C_{m+n-1}^n}.$$

补充练习 6

- 从 n 双不同的鞋子中任取 $2r(2r<n)$ 只，求下列事件发生的概率.
  (1) 没有成对的鞋子.
  (2) 只有一对鞋子.
  (3) 恰有两对鞋子.
  (4) 有 r 对鞋子.

SHANDONG UNIVERSITY 山东大学

补充练习 6

- 从 n 双不同的鞋子中任取 $2r(2r<n)$ 只，求下列事件发生的概率.
  (1) 没有成对的鞋子.
  (2) 只有一对鞋子.
  (3) 恰有两对鞋子.
  (4) 有 r 对鞋子.

$$P_1 = \frac{C_n^{2r}\cdot 2^{2r}}{C_{2n}^{2r}}.$$

$$P_2 = \frac{C_n^{1}\cdot C_{n-1}^{2r-2}\cdot 2^{2r-2}}{C_{2n}^{2r}}.$$

$$P_3 = \frac{C_n^{2}\cdot C_{n-2}^{2r-4}\cdot 2^{2r-4}}{C_{2n}^{2r}}.$$

$$P_4 = \frac{C_n^{r}}{C_{2n}^{2r}}.$$

补充练习 7

- 从装有号码 1, 2, ..., N 的球的箱子中有放回的摸了 n 次, 依次记下其号码, 试求这些号码按严格上升次序排列的概率.

补充练习 7

- 从装有号码 1, 2, ..., N 的球的箱子中有放回的摸了 n 次, 依次记下其号码, 试求这些号码按严格上升次序排列的概率.

$$P = \frac{C_N^n}{N^n}.$$

补充练习

补充练习 7

- 从装有号码 1, 2, ..., N 的球的箱子中有放回的摸了 n 次，依次记下其号码，试求这些号码按严格上升次序排列的概率.

$$P = \frac{C_N^n}{N^n}.$$

补充练习 8

- 任意从数列 1, 2, ..., N 中不放回的取出 n 个数并按大小排列成：$x_1 < x_2 < \cdots < x_m \cdots < x_n$，试求 $x_m = M$ 的概率，这里 $1 \leqslant M \leqslant N$.

补充练习

补充练习 7

- 从装有号码 1, 2, ..., N 的球的箱子中有放回的摸了 n 次，依次记下其号码，试求这些号码按严格上升次序排列的概率.

$$P = \frac{C_N^n}{N^n}.$$

补充练习 8

- 任意从数列 1, 2, ..., N 中不放回的取出 n 个数并按大小排列成：$x_1 < x_2 < \cdots < x_m \cdots < x_n$，试求 $x_m = M$ 的概率，这里 $1 \leqslant M \leqslant N$.

$$P = \frac{C_{M-1}^{m-1} \cdot C_{N-M}^{n-m}}{C_N^n}.$$

补充练习 9

- 袋中有 a 只黑球, b 只白球. 现从袋中随机地一只一只摸球 (不放回), 求第 k 次摸出黑球的概率.

补充练习 9

- 袋中有 a 只黑球, b 只白球. 现从袋中随机地一只一只摸球(不放回), 求第 k 次摸出黑球的概率.
- 摸球与顺序无关, $P=\frac{a}{a+b}$.

$$
\begin{aligned}
P(X) \quad = \quad & P(A_1A_2\cdots A_{k-1}A_k)+P(A_1A_2\cdots\overline{A_{k-1}}A_k)+\cdots \\
& +P(\overline{A_1A_2}\cdots\overline{A_{k-1}}A_k)
\end{aligned}
$$

补充练习

补充练习 9

- 袋中有 a 只黑球, b 只白球. 现从袋中随机地一只一只摸球(不放回), 求第 k 次摸出黑球的概率.
- 摸球与顺序无关, $P=\frac{a}{a+b}$.

$$\begin{aligned} P(X) &= P(A_1A_2\cdots A_{k-1}A_k)+P(A_1A_2\cdots\overline{A_{k-1}}A_k)+\cdots \\ &\quad +P(\overline{A_1A_2}\cdots\overline{A_{k-1}}A_k) \end{aligned}$$

补充练习 10

- 甲、乙均有 n 枚硬币, 全部掷完后分别计算掷出的正面数, 试求两人掷出的正面数相等的概率.

补充练习 9

- 袋中有 a 只黑球, b 只白球. 现从袋中随机地一只一只摸球(不放回), 求第 k 次摸出黑球的概率.
- 摸球与顺序无关, $P=\frac{a}{a+b}$.

$$\begin{aligned} P(X) &= P(A_1A_2\cdots A_{k-1}A_k)+P(A_1A_2\cdots\overline{A_{k-1}}A_k)+\cdots \\ &\quad +P(\overline{A_1A_2}\cdots\overline{A_{k-1}}A_k) \end{aligned}$$

补充练习 10

- 甲、乙均有 n 枚硬币, 全部掷完后分别计算掷出的正面数, 试求两人掷出的正面数相等的概率.

$$P=\sum_{k=0}^{n}B(k;n,p)^2=\sum_{k=0}^{n}\left(C_n^k\left(\frac{1}{2}\right)^n\right)^2=C_{2n}^n\left(\frac{1}{2}\right)^{2n}.$$

补充练习 11

- 选择题有 4 个答案，只有一个是正确的．不懂的学生从中随机选择．假定一个学生懂与不懂的概率都是 $\frac{1}{2}$，求答对的学生对该题确实懂的概率．

补充练习 11

- 选择题有 4 个答案, 只有一个是正确的. 不懂的学生从中随机选择. 假定一个学生懂与不懂的概率都是 $\frac{1}{2}$, 求答对的学生对该题确实懂的概率.
- $A=\{\text{懂}\}$, $P(A)=\frac{1}{2}$, $B=\{\text{答对}\}$.
- $P(B)=P(A\cap B)+P(\overline{A}\cap B)=$ $P(A)\cdot P(B|A)+P(\overline{A})\cdot P(B|\overline{A})=\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{4}$.

$$P(A|B)=\frac{P(A)\cdot P(B|A)}{P(B)}=\frac{\frac{1}{2}\cdot 1}{\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{4}}=\frac{4}{5}.$$

补充练习 12

- 假定人在一年 365 日中的任一日出生的概率是一样的，试问：在一个 730 人的单位中，至少有两人生于元旦的概率是多少？

补充练习 12

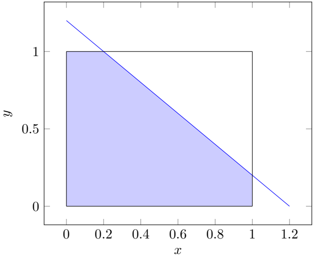
- 假定人在一年 365 日中的任一日出生的概率是一样的，试问: 在一个 730 人的单位中，至少有两人生于元旦的概率是多少?

$$
\begin{aligned}
P &= \sum_{k=2}^{730} C_{730}^{k}\left(\frac{1}{365}\right)^{k}\cdot\left(\frac{364}{365}\right)^{730-k} \\
&= 1-\left(\frac{364}{365}\right)^{730}-C_{730}^{1}\left(\frac{1}{365}\right)^{1}\cdot\left(\frac{364}{365}\right)^{729} \\
&= 0.594.
\end{aligned}
$$

补充练习 13

- 从 $(0,1)$ 中随机地取二数，求下列概率:
  (1) 两数之和小于 1.2.
  (2) 两数之积小于 $\frac{1}{4}$.
  (3) 以上两个要求同时满足.
  (4) 记两数为 b 和 c，求方程 $x^2+bx+c=0$ 有实根的概率.

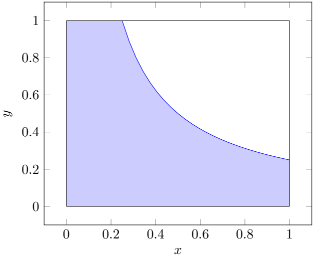
(1) $x+y<1.2$, $P=1-\frac{1}{2}\cdot 0.8\cdot 0.8=0.68$.



补充练习 13

- 从 $(0,1)$ 中随机地取二数, 求下列概率:
  (1) 两数之和小于 1.2.
  (2) 两数之积小于 $\frac{1}{4}$.
  (3) 以上两个要求同时满足.
  (4) 记两数为 b 和 c, 求方程 $x^2+bx+c=0$ 有实根的概率.

(2) $x\cdot y<\frac{1}{4}$, $P=0.25+\int_{0.25}^{1}\frac{1}{4x}\,dx=0.5966$.

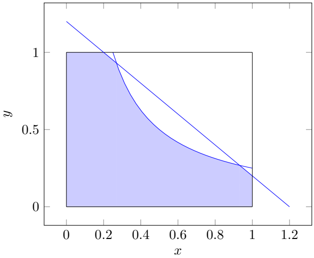


补充练习 13

- 从 $(0,1)$ 中随机地取二数，求下列概率:
  (1) 两数之和小于 1.2.
  (2) 两数之积小于 $\frac{1}{4}$.
  (3) 以上两个要求同时满足.
  (4) 记两数为 b 和 c，求方程 $x^2+bx+c=0$ 有实根的概率.

(3) $x+y=1.2,\ x\cdot y=\frac{1}{4} \Rightarrow x_1=0.2683,\ x_2=0.9317.$

- $P=0.2+\int_{0.2}^{0.2683}(1.2-x)\,dx+\int_{0.2683}^{0.9317}\frac{1}{4x}\,dx+\int_{0.9317}^{1}(1.2-x)\,dx=0.593.$



补充练习 13

- 从 $(0,1)$ 中随机地取二数，求下列概率：
  (1) 两数之和小于 1.2.
  (2) 两数之积小于 $\frac{1}{4}$.
  (3) 以上两个要求同时满足.
  (4) 记两数为 b 和 c，求方程 $x^2+bx+c=0$ 有实根的概率.

(4) $b^2-4\cdot c\geqslant 0$，$c\leqslant\frac{b^2}{4}$.

- $P=\int_0^1\frac{b^2}{4}\,db=\frac{1}{12}$.