随机事件 (Random Event)

September 5, 2024

确定现象和随机现象

- **必然现象:** 条件完全决定结果
  - 太阳东升西落
  - 可导函数必然连续
  - 树叶的形状
- **随机现象:** 条件不能完全决定结果
  - 生男孩、生女孩
  - 没有两片树叶是完全相同的
  - 抛投硬币: 字还是花朝上

随机现象的统计规律性

由于随机现象中, 条件不能完全决定结果, 所以

- 条件和结果之间没有确定性的联系
- 随机现象在单次的试验/观察中得到结果具有偶然性, 无法准确预知会观察到哪一种结果
  - 抛投硬币: 字还是花朝上

随机现象是否无规律可言呢? **未必**

- 大量重复试验或者观察中, 出现的结果具有一定的"统计"规律性

概率: 研究随机现象统计规律性的学科

随机试验: 例 0.6

回到 Poker Hands 的例子:

- **Gaming**: a one-pair hand consists of 5 cards
- **Question 1**: the probability of “the count of a one-pair hand is less than point 12”
- **Using sampling-based methods**:
  1. build a trial by sampling 5 cards from 52 cards – 试验, 不确定性
  2. repeat n trials – 可重复, 相同条件下多次试验
  3. count the number m of “appropriate” trials – 多结果
  4. regard the frequency $\frac{m}{n}$ as the probability – 具有“统计”规律性

随机试验

随机试验:

- 多结果
- 不确定
- 可重复
- 具有“统计”规律性

大量重复试验或者观察中, 出现的结果具有一定的“统计”规律性

样本点、样本空间

- **样本点**: 随机试验中的每一种可能的结果
  - sampling $\{2H, 2S, 5C, 7S, 10D\} \stackrel{\text{def}}{=} \omega$
- **样本空间**: 所有样本点组成的集合, 记为 $\Omega = \{\omega\}$
  - 势 $|\Omega|$ 表示样本空间的大小, 即样本点的个数
- **样本空间的类型**:
  - 有限样本空间：样本点个数是有限的
    - Poker Hands: $|\Omega| = \binom{52}{5} = 2,598,960$
  - 无限可列样本空间: 样本点个数是无限的、但可列的
    - 新生人口数量
  - 不可列样本空间: 样本点个数是无限的、且不可列的
    - 随机取一个实数是无理数

随机事件

样本点的集合, 记为 A, B, C

- Poker Hand “count less than point 12”
  - $A = \{\omega_1, \omega_2, \omega_3, \omega_4\}$
  - $\omega_1 = \{2S, 2H, 2C, 2D, 3S\}$
  - $\omega_2 = \{2S, 2H, 2C, 2D, 3H\}$
  - $\omega_3 = \{2S, 2H, 2C, 2D, 3C\}$
  - $\omega_4 = \{2S, 2H, 2C, 2D, 3D\}$
- 随机事件是样本空间的子集
- 事件 A 发生 if and only if 试验的结果 ω 是集合 A 中的元素

随机事件

- 基本事件: 只包含一样本点的事件;
- 复合事件: 只包含两个或两个以上样本点的事件;
- 必然事件: 样本空间 Ω 包含所有样本点, 每次试验必然发生;
- 不可能事件: 某事件在每次试验中都不发生, 事件不包含任何样本点, 用空集 $\emptyset$ 表示.

例 0.7 随机试验 E: 抛掷一枚骰子观察其出现的点数, 其样本空间 $\Omega = \{1, 2, \ldots, 6\}$. 判断如下事件的类型

- 事件 A 表示抛骰子的点数为 2
- 事件 B 表示抛骰子的点数为偶数
- 事件 C 表示抛骰子的点数大于 7
- 事件 D 表示抛骰子的点数小于 7

样本空间与随机事件：例 0.7

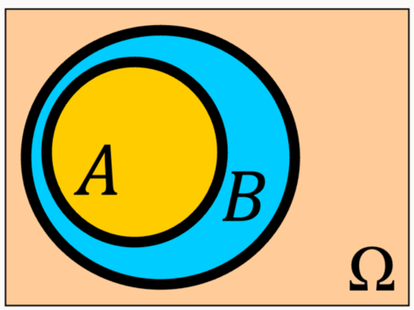
题目: 随机试验 E: 抛掷一枚骰子观察其出现的点数, 其样本空间 $\Omega = \{1, 2, \ldots, 6\}$.

则有:

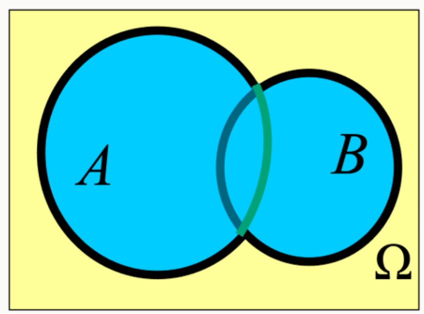
- 事件 A 表示抛骰子的点数为 2, 则 $A = \{2\}$ 为基本事件;
- 事件 B 表示抛骰子的点数为偶数, 则 $B = \{2, 4, 6\}$ 为复合事件;
- 事件 C 表示抛骰子的点数大于 7, 则 $C = \emptyset$ 为不可能事件;
- 事件 D 表示抛骰子的点数小于 7, 则 $D = \Omega$ 为必然事件.

事件之间的关系

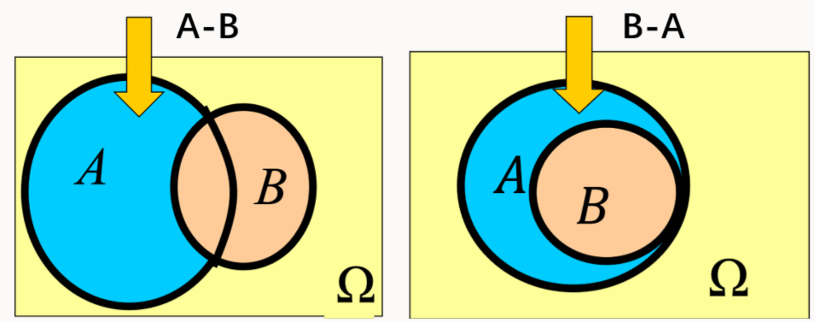
- **包含**: 事件 A 的发生必然导致事件 B 的发生, 记 $A \subseteq B$
  - $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$
- **并/和**: 事件 A 和事件 B 至少发生一个, 记为 $A \cup B$
- **差**: 事件 A 发生而事件 B 不发生, 记为 $A - B$



包含 (inclusion)



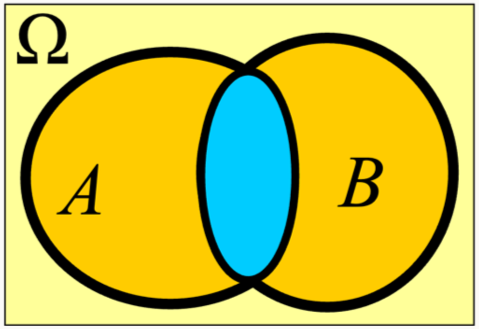
并 (union or sum)



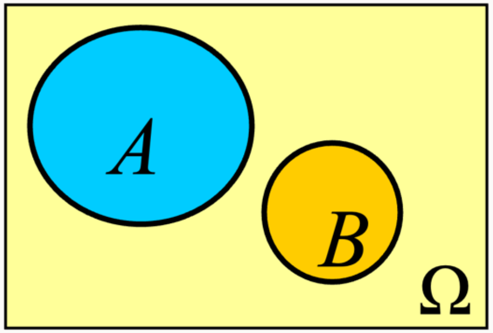
差 (union or sum)

事件之间的关系

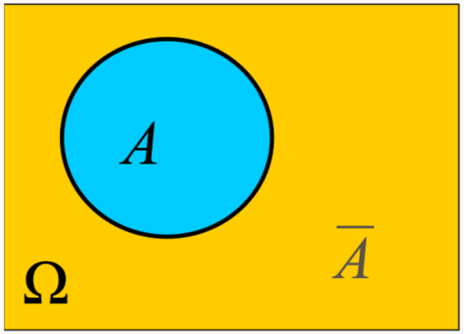
- **交**: 事件 A 和事件 B 同时发生, 记为 $A \cap B$ or AB
- **互斥**: 事件 A 和事件 B 必不可能同时发生, 记为 $A \cap B = \emptyset$
- **对立**: 事件 A 不发生的事件, 记为 $\bar{A}$
  - $A \cap \bar{A} = \emptyset$ and $A \cup \bar{A} = \Omega$



交 (intersect)



互斥 (mutual exclusion)



对立 (opposite)

概率论与集合论之间的对应关系

记号	概率论	集合论
$\bar{A}$	A 的对立事件	A 的补集
$A \subset B$	A 发生必然导致 B 发生	A 是 B 的子集
$A = B$	事件 A 与事件 B 相等	集合 A 和集合 B 相等
$A \cup B$	事件 A 与事件 B 的和	集合 A 和集合 B 并集
$A \cap B$	事件 A 与事件 B 的积	集合 A 和集合 B 交集
$A - B$	事件 A 与事件 B 的差	集合 A 和集合 B 差集
$A \cap B = \emptyset$	事件 A 与事件 B 互不相容	集合 A 和集合 B 没有相同的元素

事件的运算规律——遵循集合论运算规律

- 幂等律: $A \cup A = A$, $A \cap A = A$
- 交换律: $A \cup B = B \cup A$, $A \cap B = B \cap A$
- 结合律:
  - $(A \cup B) \cup C = A \cup (B \cup C)$
  - $(A \cap B) \cap C = A \cap (B \cap C)$
- 分配律:
  - $A \cap (B \cup C) \cup C = (A \cap B) \cup (A \cap C)$
  - $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$
- 对偶律: $\overline{A \cup B} = \bar{A}\bar{B}$, $\overline{AB} = \bar{A} \cup \bar{B}$

随机事件: 例题 0.8

例 0.8 设 A, B, C 为三个随机事件, 则有:

- 事件 A 与 B 同时发生, 而事件 C 不发生的事件可表示为 $AB\bar{C}$ 或 $AB - C$;
- 三个事件中至少有一个发生的事件可表示为 $A \cup B \cup C$;
- 三个事件中恰好有一个发生的事件可表示为 $(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C)$;
- 三个事件中至多有一个发生的事件可表示为 $(A\bar{B}\bar{C}) \cup (\bar{A}B\bar{C}) \cup (\bar{A}\bar{B}C) \cup (\bar{A}\bar{B}\bar{C})$ 或 $\overline{AB \cup AC \cup BC}$
- 三个事件中至少有两个发生的事件可表示为 $AB \cup AC \cup BC$;
- 三个事件中至多有两个发生的事件可表示为 $\bar{A} \cup \bar{B} \cup \bar{C}$;
- 三个事件中恰好有两个发生的事件 $(AB\bar{C}) \cup (AC\bar{B}) \cup (BC\bar{A})$.

随机事件: 例题 0.9

例 0.9 设 A, B, C 为三个随机事件, 求证:

- $(\bar{A} \cup B)(A \cup B)(\bar{A} \cup \bar{B})(A \cup \bar{B}) = \emptyset$
- $(A - B) \cup (B - C) = (A \cup B) - BC$

解答：例 0.9

题目: 设 A, B, C 为三个随机事件, 求证:

$$(\bar{A} \cup B)(A \cup B)(\bar{A} \cup \bar{B})(A \cup \bar{B}) = \emptyset \qquad \text{和} \qquad (A - B) \cup (B - C) = (A \cup B) - BC$$

解答:

- 根据事件的分配律有 $(\bar{A} \cup B)(A \cup B) = (A \cap \bar{A}) \cup B = B$ 以及 $(\bar{A} \cup \bar{B})(A \cup \bar{B}) = \bar{B}$, 由此可得 $(\bar{A} \cup B)(A \cup B)(\bar{A} \cup \bar{B})(A \cup \bar{B}) = B \cap \bar{B} = \emptyset$.
- 根据事件的差 $A - B = A\bar{B}$ 可得 $(A - B) \cup (B - C) = (A\bar{B}) \cup (B\bar{C})$. 根据事件的分配律和德摩根律有

$$\begin{aligned}(A \cup B) - BC &= (A \cup B)\overline{BC} = (A \cup B) \cap (\bar{B} \cup \bar{C}) \\ &= (A\bar{B}) \cup (A\bar{C}) \cup (B\bar{B}) \cup (B\bar{C}) = (A\bar{B}) \cup (A\bar{C}) \cup (B\bar{C}).\end{aligned}$$

  由此可得 $((A - B) \cup (B - C)) \subset ((A \cup B) - BC)$. 只需进一步证明 $A\bar{C} \subset (A\bar{B}) \subset (B\bar{C})$, 对任意 $x \in A\bar{C}$, 有 $x \in A$ 或 $x \in \bar{C}$, 再根据 $x \in B$ 或 $x \in \bar{B}$ 或 $x \in A\bar{B}$ 或 $x \in B\bar{C}$ 成立.

概率与频率 (Probability and Frequency)

September 5, 2024

回忆: 例 0.6 的解法一

回到 Poker Hands 的例子:

- **Gaming**: a one-pair hand consists of 5 cards
- **Question 1**: the probability of “the count of a one-pair hand is less than point 12”
- **Using sampling-based methods**:

1. build a trial by sampling 5 cards from 52 cards – 试验, 不确定性
2. repeat n trials – 可重复, 相同条件下多次试验
3. count the number m of “appropriate” trials – 多结果
4. regard the frequency $\frac{m}{n}$ as the probability – 具有“统计”规律性

频率

定义 0.1 在相同的条件下, 进行了 n 次试验, n 次试验中事件 A 发生的次数为 m, 则称事件 A 发生的频率为:

$$f_n(A) = \frac{m}{n}$$

事件的频率在一定程度上反映了事件发生的可能性, 其性质包括:

- $0 \leq f_n(A) \leq 1$
- $f_n(\Omega) = 1$
- 互斥事件的频率可加:

$$f_n(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} f_n(A_i)$$

频率的稳定性

频率, 在试验中, 表现出随机性

在大量的重复试验中, 事件的频率通常在一个常数 p 附近摆动, 且随着试验次数的增大, 摆动幅度会越来越小, 这种规律称为 **频率的稳定性**.

频率的稳定性, 即之前提到的 "统计规律性"

- 是随机事件本身固有的客观属性
- 是 sampling-based methods / 统计试验设计的基础 (cornerstone)
- 可用于度量随机事件发生的可能性大小

频率的稳定性——投币实验

历史上多人对重复投掷硬币的试验, 下面是试验统计结果:

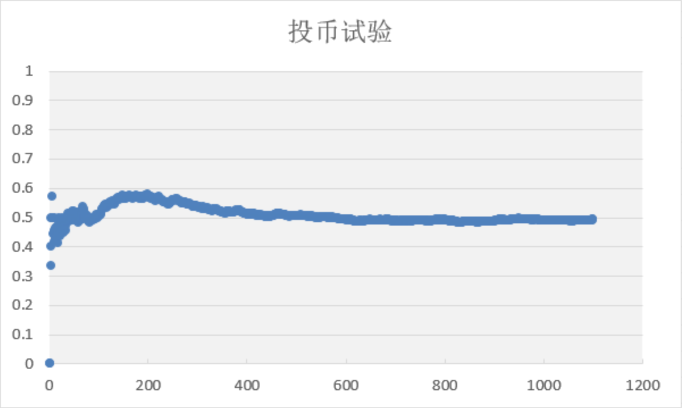
Experimenters	n	n **of head**	**frequency of head**
德 · 摩根	2048	1061	0.5181
蒲丰	4040	2048	0.5069
K· 皮尔逊	12000	6019	0.5016
K· 皮尔逊	24000	12012	0.5005

频率的稳定性——计算机模拟

抛硬币频率趋势曲线

		1000	1100	1200	1300	1400	1500	1600	1700	1800	1900	2000
试验次数	1100											

序号	正反情况	正面频率
1	0	0
2	0	0
3	1	0.333333
4	1	0.5
5	0	0.4
6	1	0.5
7	1	0.571429
8	0	0.5
9	0	0.444444
10	1	0.5
11	0	0.454545
12	0	0.416667
13	1	0.461538
14	0	0.428571
15	1	0.466667
16	0	0.4375
17	0	0.411765
18	1	0.444444
19	1	0.473684
20	1	0.5
21	0	0.47619
22	0	0.454545
23	0	0.434783
24	1	0.458333



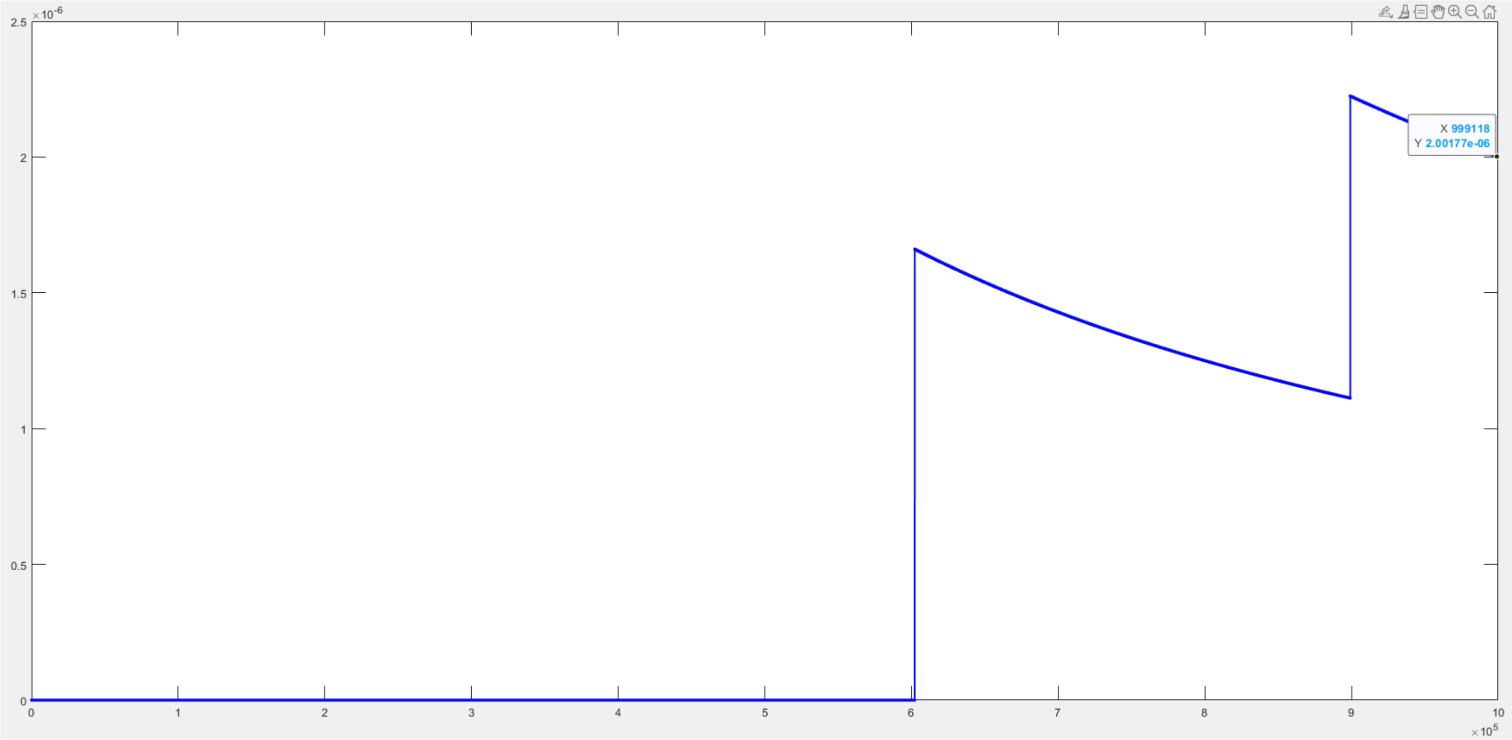
试验次数较少时, 频率有波动性; 试验次数较大时, 频率具有稳定性.

回忆: 例 0.6 的解法二

回到 Poker Hands 的例子:

- **Gaming**: a one-pair hand consists of 5 cards
- **Question 1**: the probability of “the count of a one-pair hand is less than point 12”
- **Using combination-based methods:**
  1. the target one-pair hand comprises $\{2, 2, 2, 2, 3\}$
  2. possible counts 4
  3. all combination $\binom{52}{5} = 2,598,960$
  4. the probability $\frac{4}{2{,}598{,}960}$

模拟实验 for 例 0.6



The figure plots the curve of $\frac{m}{n}$. It is observed that

$$\frac{m}{n} \to \frac{4}{2,598,960}, \quad \text{that is,} \quad \lim_{n\to\infty} \frac{m}{n} = \frac{4}{2,598,960}$$

概率的统计学定义

定义 0.2 随机事件 A 在大量重复试验中发生的频率总是稳定地在一个常数 p 附近摆动, 随着试验次数的增加而摆幅逐渐越小, 则称常数 p 为事件 A 发生的概率, 记为 $P(A) = p$.

事件的概率用于度量事件发生的可能性, 在一定程度上反映了围绕事件的随机试验的统计学规律, 其性质包括:

- $0 \le P(A) \le 1$
- $P(\Omega) = 1$
- 互斥事件的频率可加:

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{i=1}^{k} P(A_i)$$

概率 VS 频率

概率:

- 用于度量事件发生的可能性
- 客观上是恒定的

频率:

- 用在一定程度上反映了事件发生的可能性
- 在试验中具有随机性

频率是概率的一个近似

- 因为若试验次数足够多时, 频率与概率非常接近
- 因此定义 0.1 和定义 0.2 中三点性质相似

注意: 概率的统计定义存在数学上的不严谨性, 在实际中几乎不可能每一个事件做大量重复的试验来计算频率, 进而近似概率受到频率的稳定性及其性质的启发给出严谨的概率公理化体系.

概率的公理化定义

苏联数学家柯尔莫哥洛夫 (Kolmogorov) 于 1933 年给出了概率的公理化定义, 通过规定概率具备的基本性质来定义概率

定义 0.3 概率公理化在可测空间 (Ω, Σ) 上，若函数 $P : \Sigma \to [0, 1]$ 满足下列条件:

- $0 \leq P(A) \leq 1$
- $P(\Omega) = 1$
- (可列的) 互斥事件的概率可加 $P(\cup_{i=1}^{\infty} A_i) = \cup_{i=1}^{\infty} P(A_i)$

称 $P(A)$ 为事件 A 的概率, (Ω, Σ, P) 为概率空间.

简明扼要地刻画了概率的定义, 为现代概率论奠定了基础, 是概率论发展历史上的一个里程碑, 从此被公认为数学的一个分支, 这也是概率论统计被评定为一级学科的原因之一 (与数学并列).

概率的计算性质

- 对于不可能发生的事件 $\emptyset$, 有 $P(\emptyset) = 0$
- 对于必然发生的事件 Ω, 有 $P(\Omega) = 1$
- 补运算 $P(A) = 1 - P(\bar{A})$
- 若 $B \subseteq A$, 有 $P(A - B) = P(A) - P(B)$ 和 $P(A) \geq P(B)$
- 对于任意随机事件 A 和 B, 有

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

容斥原理 (Inclusion-Exclusion Principle)

- 对于任意两个随机事件 A 和 B, 有

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

- 对于任意三个随机事件 A, B, C, 有

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC)$$

- 对于任意多个随机事件 $A_1, A_2, \ldots, A_n$, 有

$$P(\cup_{i=1}^{n} A_i) = \sum_{i}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) + \cdots + (-1)^{n-1} P(A_1 \ldots A_n)$$

可进一步简化为

$$P(\cup_{i=1}^{n} A_i) = \sum_{r=1}^{n} (-1)^{r+1} \left(\sum_{i_1 < \cdots < i_r} P(A_{i_1} \ldots A_{i_r}) \right)$$

不等式

- Union Bounds: 对于任意多个随机事件 $A_1, A_2, \ldots, A_n$, 有

$$P(\cup_{i=1}^{n} A_i) \leq P(A_1) + P(A_2) + \cdots + P(A_n)$$

- Bonferroni Inequalities: 对于任意多个随机事件 $A_1, A_2, \ldots, A_n$, 有

$$P(\cup_{i=1}^{n} A_i) \leq \cup_{i=1}^{n} P(A_i)$$

$$P(\cup_{i=1}^{n} A_i) \geq \cup_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j)$$

$$P(\cup_{i=1}^{n} A_i) \leq \cup_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i A_j) + \sum_{i<j<k} P(A_i A_j A_k)$$

该结论可以依次类推

概率与频率: 例题 0.10

例 0.10 设 $P(A) = p$, $P(B) = q$, $P(AB) = r$, 请用 p, q, r 来表示下列事件的概率

- $P(\bar{A} \cup \bar{B})$
- $P(\bar{A}B)$
- $P(\bar{A} \cup B)$
- $P(\bar{A} \cap \bar{B})$

解答：例 0.10

- 对于问题 1), 根据事件的德摩根律有

$$P(\bar{A} \cup \bar{B}) = P(\overline{AB}) = 1 - r.$$

- 对于问题 2), 根据差事件的定义有

$$P(\bar{A}B) = P(B - A) = P(B) - P(AB) = q - r.$$

- 对于问题 3), 根据容斥原理有

$$P(\bar{A} \cup B) = P(\bar{A}) + P(B) - P(\bar{A}B) = 1 - p + q - (q - r) = 1 - p + r.$$

- 对于问题 4), 根据德摩根律与容斥原理有

$$P(\bar{A} \cup \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 + r - p - q.$$

概率与频率: 例题 0.11

例 0.11 设三个随机事件 A, B, C 满足

$$P(A) = P(B) = P(C) = 1/4$$
$$P(AB) = 0$$
$$P(AC) = P(BC) = 1/16$$

求事件 A, B, C 中至少有一个事件发生的概率.

解答：例 0.11

题目: 设三个随机事件 A, B, C 满足 $P(A) = P(B) = P(C) = 1/4$, $P(AB) = 0$, $P(AC) = P(BC) = 1/16$, 求事件 A, B, C 中至少有一个事件发生的概率.

解答:

- 根据三个事件的容斥原理有

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(AB) - P(AC) - P(BC) + P(ABC) \\ &= 3/4 - 1/8 + P(ABC) \end{aligned}$$

- 根据 $P(AB) = 0$ 和 $ABC \subset AB$ 可知

$$0 \leq P(ABC) \leq P(AB) = 0$$

由此可知事件 A, B, C 中至少有一个事件发生的概率为 5/8.