第一章　随机事件与概率

在自然界和人类社会中存在两类不同的现象, 一类是确定性现象, 另一类是非确定性现象, 也叫随机现象. 概率论是研究随机现象统计规律性的科学, 其任务是对随机现象发生的可能性给出度量方式, 避免" 直觉陷阱".

例如, 在篮球运动中通常会认为一名运动员如果连续命中多球, 那么下一次投篮命中率也将会提高, 这被称为" 手热效应". 然而实际研究显示多数情况下是独立的. 再如, 在 COVID-19 疫情初期, 由于确诊患者中有许多患者与武汉有过交集, 于是社会上形成了" 与武汉有过交集的人都有很大的概率感染 COVID-19" 的舆论. 实际上, 数据表明, 即便进行一次核酸检测呈阳性的武汉居民, 其的确感染 COVID-19 的概率也低于 70%. 逻辑推理与数据分析, 是对随机现象发生的可能性进行分析的科学手段.

本章, 我们学习概率论中的一些基本概念, 为后面的学习做好铺垫.

1.1　随机事件和样本空间

在概率论中, 随机现象描述为随机试验. 在一次试验 E 中, 如果该试验的结果不能事先确定, 但所有可能的结果是已知的, 则该试验称为**随机试验**. 例如,

- $E_1 =$" 掷一枚骰子, 观察其点数";

- $E_2 =$" 先后掷两枚硬币, 观察正反面";

- $E_3 =$" 观察某电子元件损坏时已持续运行的时间".

随机试验可能的结果称为该试验的**样本点**, 常记为 $\omega_1, \omega_2, \cdots$. 样本点的全体称为**样本空间**, 常记为 Ω. 例如, 上述随机试验的样本空间分别为

- $\Omega_1 = \{1, 2, 3, 4, 5, 6\}$, 含有 6 个样本点;

- $\Omega_2 = \{(H, T), (H, H), (T, H), (T, T)\} \quad (H(head) = \text{正面}, T(tail) = \text{反面})$;

- $\Omega_3 = \{t : t \geqslant 0\} = [0, \infty)$, 有无穷多的样本点.

样本空间 Ω 的每个子集称为随机试验的一个**随机事件**, 简称**事件**, 常用大写字母 A, B, C, $\cdots$, 表示. 若某次随机试验的结果 ω 落于某事件 A 中, 即 $\omega \in A$, 则称事件 A 在该次试验中**发生**.

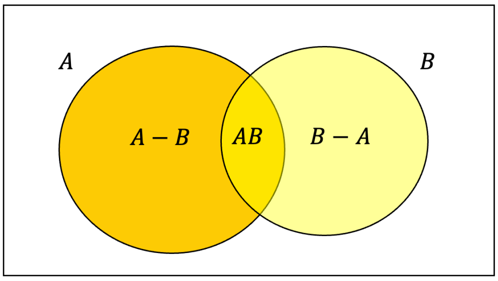

图 1.1: 随机事件的运算图示

例如, 在 E_1 中, $A = \{6\}$, $B = \{2, 4, 6\} = \{$偶数点$\}$ 均为随机事件. 若在一次试验中掷骰子掷出 6 点, 则这两个事件都发生. 此外, 事件 A 只包含一个样本点, 这样的事件称为**基本事件**; 而事件 B 包含不同的基本事件, 称为**复合事件**. 再如, 在 E_3 中, $A = \{t : t \geqslant 500\}$ (或解读为合格) 也是随机事件.

特别地, 样本空间 Ω 的平凡子集 Ω 称为**必然事件**, 空集 $\emptyset$ 称为**不可能事件**. 注意, 不可能事件不等同于概率为零的事件.

例 1.1. 某士兵向一标靶连续射击三次, 观察这三次射击的命中情况. 写出该随机试验的样本空间 Ω 以及事件 $A =$" 标靶被命中" 所包含的样本点.

解. 记 1 = 命中, 0 = 没命中, 则样本空间 $\Omega = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_i = 1$ 或 $0\}$,

$$A = \left\{(x_1, x_2, x_3) \in \Omega : x_1^2 + x_2^2 + x_3^2 \neq 0\right\}. \qquad \square$$

1.2 随机事件的关系和运算

由于随机事件即为样本空间的子集, 它们的关系和运算与集合间的关系和运算是相通的. 例如, $\overline{A} := \Omega \setminus A$ 称为事件 A 的**对立事件**或**逆事件**. 易见, 任何事件与其对立事件必有一个会发生, 但不会同时发生. 再如,

(i) $A \subset B$ 表示 A 发生则 B 必然发生;

(ii) $A = B$ 等价于 $A \subset B$ 且 $B \subset A$, 表示两事件相等或等价;

(iii) $A \cup B$ 发生等价于 A 发生或 B 发生 (即至少有一个发生);

(iv) $A \cap B$ (或 AB) 发生等价于 A 和 B 同时发生;

(v) $A - B = A\overline{B}$ 发生等价于 A 发生但 B 不发生, 见图1.1.

此外, 当事件 $AB = \emptyset$ 时, 称事件 A, B **互不相容**. 此时, $A \cup B$ 常记为 $A + B$, 即 $A + B = A \cup B$ 且 $AB = \emptyset$.

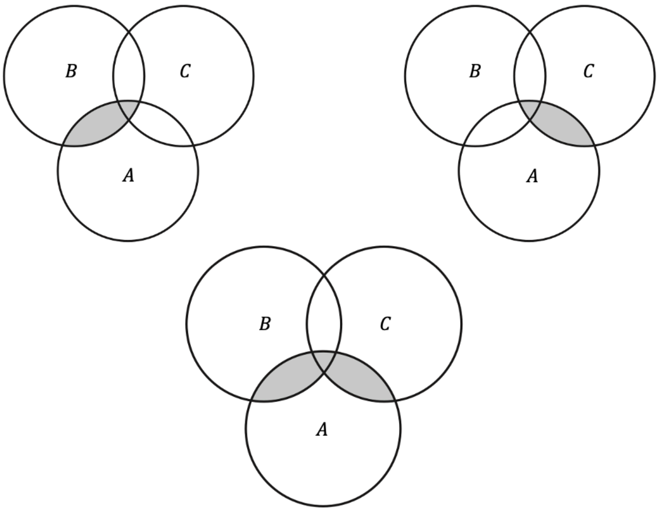

图 1.2: 分配率 $A(B \cup C) = (AB) \cup (AC)$.

命题 1.2. 随机事件满足如下运算律:

(i) 交换律: $A \cup B = B \cup A$, $A \cap B = B \cap A$;

(ii) 结合律: $(A \cup B) \cup C = A \cup (B \cup C)$, $(AB)C = A(BC)$;

(iii) 分配率: $A(B \cup C) = (AB) \cup (AC)$, $A \cup (BC) = (A \cup B)(A \cup C)$;

(iv) 对偶律 (De Morgan 律): $\overline{A \cup B} = \overline{A} \cap \overline{B}$, $\overline{A \cap B} = \overline{A} \cup \overline{B}$,

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i}, \quad \overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i}.$$

这些运算律与集合间的远算律相吻合, 因为事件本身就可解读为集合. 例如, 若事件 $\overline{A \cup B}$ 发生, 即若某样本 $\omega \in \overline{A \cup B}$, 则根据集合的关系必有 $\omega \in \overline{A} \cap \overline{B}$, 即 $\overline{A} \cap \overline{B}$ 必发生. 因此, 有关系 $\overline{A \cup B} \subset \overline{A} \cap \overline{B}$. 同理可证反向包含, 于是有对偶律 $\overline{A \cup B} = \overline{A} \cap \overline{B}$.

这些运算律也可通过画图的方式帮助理解, 如图1.2.

例 1.3. 某士兵向一标靶连续射击三次, 观察这三次射击的命中情况. 记 A_i 为 "第 i 次命中标靶 " $(i = 1, 2, 3)$, 试用 A_1, A_2, A_3 及其运算式表示下面事件:

(i) B_1 = "前两次至少命中一次";

(ii) B_2 = "只有第一次未命中";

(iii) B_3 = "第一次命中且后两次至少命中一次".

解. $B_1 = A_1 \cup A_2$; $B_2 = \overline{A}_1 A_2 A_3$; $B_3 = A_1(A_2 \cup A_3)$. □

在上例中, 事件 B_1 与第三次射击的结果 A_3 无关. 换言之, B_1 发生的概率与" 该士兵向标靶连续射击两次至少命中一次" 发生的概率是相同的, 尽管这两个事件来自不同的试验 (一个是连续射击三次的试验, 一个是连续射击两次的试验, 概率空间也不相同).

1.3 随机事件的概率及其计算

现在讨论随机事件发生的" 可能性大小" 的问题. 这种可能性可由两种方式来定量分析, 一是通过试验或统计手段得到该事件发生的" 相对频率", 二是通过理论分析获得其发生的" 概率".

频率与概率

定义 1.4. 设某随机试验在相同条件下重复进行了 N 次, 而随机事件 A 发生了 n_A 次, 则称比值

$$f_N(A) = \frac{n_A}{N}$$

为随机事件 A 发生的**相对频率**, 简称**频率**.

若事件 A 在重复试验中发生的频率越高, 那么可以合理地认为它在该 (单次) 随机试验中发生的可能性就越大.

不难理解, 事件的频率与重复试验的次数 N 有关. 大量试验证实, 当重复试验的次数 N 逐渐增大时, 频率 $f_N(A)$ 呈现出稳定性, 逐渐稳定于某个常数. 这种"频率稳定性"即通常所说的统计规律性. 例如, 历史上有数学家做过" 抛硬币" 试验, 正面向上的相对频率统计如表1.1.

学者	试验次数 N	正面向上次数 n_A	相对频率 $f_N(A)$
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

表 1.1: 抛硬币试验的统计数据

因此, 一个直观的想法是通过对频率取极限以获得刻画事件 A 发生的可能性的量, 即概率:

$$P(A) = \lim_{N\to\infty} f_N(A).$$

概率的这种定义方式是直观的, 也具有相当的合理性 (事实上我们将看到, 当 $N \to \infty$ 时, 频率 $f_N(A)$ 的确在一定意义下接近概率 $P(A)$). 然而在数学上, 它却有一个重大缺陷: 如何知道或判断某重复试验所获得的频率序列的确趋向某个常数, 而且进行另一次重复试验

它会收敛到相同的常数? 即便将频率趋向常数这件事当做研究随机现象的前提 (或公理) 提出, 在实际问题中, 也不可能对每一个事件都做大量的试验以获得事件发生的频率序列.

现代概率论的公理化方法基于下面三个更为合理的假设.

定义 1.5 (概率的公理化定义, Kolmogorov (1933))**.** 设 Ω 是随机试验 E 的样本空间. 若对每个事件 $A \subset \Omega$ 都定义有一个数 $P(A)$, 满足

(i) 非负性: 对任意事件 A, $P(A) \geqslant 0$;

(ii) 规范性: $P(\Omega) = 1$;

(iii) 可列可加性: 设 $A_1, A_2, \cdots$ 是两两互不相容的事件 (即 $A_iA_j = \emptyset$, $i \neq j$), 则

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots,$$

则称 $P(A)$ 为事件 A 的**概率**.

我们后面会看到, 如此定义的事件 A 的概率即为其频率在某种意义下的极限.

概率的性质

由概率的定义可推得一些重要性质.

命题 1.6. 不可能事件的概率为零, 即 $P(\emptyset) = 0$.

证明. 取 $A_i = \emptyset$, $i = 1, 2, \cdots$, 则由可列可加性,

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

因此, 由概率的非负性, $P(\emptyset) = 0$. □

命题 1.7 (有限可加性)**.** 对互不相容的事件 $A_1, A_2, \cdots, A_n$, 有

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

证明. 取 $A_i = \emptyset$, $i = n+1, n+2, \cdots$, 则由可列可加性,

$$\begin{aligned} P\left(\bigcup_{i=1}^{n} A_i\right) &= P\left(\bigcup_{i=1}^{\infty} A_i\right) \\ &= \sum_{i=1}^{n} P(A_i) + \sum_{i=n+1}^{\infty} P(\emptyset) \\ &= \sum_{i=1}^{n} P(A_i). \end{aligned}$$

□

命题 1.8 (逆事件概率)**.** $P(\overline{A}) = 1 - P(A)$.

证明. 由于 $\Omega = A + \overline{A}$, 由有限可加性和规范性, 有

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

□

命题 1.9 (单调性)**.** 若 A, B 是两个事件, $A \subset B$, 则

$$P(B - A) = P(B) - P(A),$$
$$P(A) \leqslant P(B).$$

证明. 由于 $A \subset B$, 有 $B = B - A + A$. 因此, 由有限可加性,

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

又 $P(B - A) \geqslant 0$, 有

$$P(B) \geqslant P(A).$$

□

命题 1.10. 对任意事件 A, 有

$$0 \leqslant P(A) \leqslant 1.$$

证明. 由正定性, 单调性和规范性, 有

$$0 \leqslant P(A) \leqslant P(\Omega) = 1.$$

□

命题 1.11 (加法公式)**.** 对任意两事件 A, B, 有

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

证明. 由于 $A \cup B = A + (B - AB)$,

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

□

此加法公式可轻易推广至多个事件的情形:

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leqslant i < j \leqslant n} P(A_i A_j) + \sum_{1 \leqslant i < j < k \leqslant n} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 A_2 \cdots A_n).$$

下面我们说明概率作为一个映射, 关于事件具有一定的" 连续性".

设 $\{A_k\}_{k \in \mathbb{N}}$ 是一个单调递增事件列, 即满足 $A_i \subset A_j$ $(i < j)$, 则 $\cup_{k \in \mathbb{N}} A_k$ 是一个事件, 称为事件列 $\{A_k\}_{k \in \mathbb{N}}$ 的**极限**, 记为

$$\lim_{k \to \infty} A_k = \bigcup_{i \in \mathbb{N}} A_i.$$

同理, 对单调递减的事件列 $\{A_k\}_{k \in \mathbb{N}}$, 定义其**极限**为

$$\lim_{k \to \infty} A_k = \bigcap_{i \in \mathbb{N}} A_i.$$

命题 1.12 (概率的连续性)**.** 对于单调 (递增或递减) 的事件列 $\{A_k\}_{k\in\mathbb{N}}$, 有

$$\lim_{k\to\infty} P(A_k) = P\left(\lim_{k\to\infty} A_k\right) \qquad (\text{即 “}A_k \to A\text{” } \Rightarrow \text{ “}P(A_k) \to P(A)\text{”}).$$

证明. 只证 $\{A_k\}_{k\in\mathbb{N}}$ 递增的情形. 定义事件列 $\{B_k\}_{k\in\mathbb{N}}$ 如下:

$$\begin{aligned} B_1 &:= A_1, \\ B_k &:= A_k \bigcap \overline{\cup_{i=1}^{k-1} A_j} = A_k \overline{A}_{k-1}, \quad k > 1. \end{aligned}$$

这里用到了 $\{A_k\}_{k\in\mathbb{N}}$ 的递增性. 易见, 事件列 $\{B_k\}_{k\in\mathbb{N}}$ 互不相容且

$$\bigcup_{k\in\mathbb{N}} A_k = \sum_{k\in\mathbb{N}} B_k, \quad \sum_{k=1}^{n} B_k = A_n \quad (n \geqslant 1).$$

于是, 由概率的可数/有限可加性,

$$P\left(\bigcup_{k\in\mathbb{N}} A_k\right) = P\left(\sum_{k\in\mathbb{N}} B_k\right) = \sum_{k\in\mathbb{N}} P(B_k) = \lim_{n\to\infty} \sum_{k=1}^{n} P(B_k) = \lim_{n\to\infty} P\left(\sum_{k=1}^{n} B_k\right) = \lim_{n\to\infty} P(A_n). \qquad \square$$

1.4　古典概型与几何概型

本节讨论古典概型与几何概型. 这两种随机试验的共同特点是试验结果在可能的取值范围内等概率分布.

1.4.1　古典概型

古典概型是指这样的随机试验:

1. 有限性: 样本空间只包含有限个样本点;
2. 等可能性: 这有限个样本点的发生是等可能的.

这样的试验是大量存在的, 而且在概率论发展的初期曾是主要的研究对象, 因此称为古典概型.

设某古典概型的样本空间包含 n 个样本点, 即 $\Omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$, 则由定义,

$$P(\omega_i) = \frac{1}{n}, \quad i = 1, 2, \cdots, n.$$

此时, 任意事件 $A \subset \Omega$ 发生的概率

$$\boxed{P(A) = \frac{\text{事件 } A \text{ 包含的样本点数 } \sharp A}{\text{样本空间 } \Omega \text{ 包含的样本点数 } \sharp\Omega}.} \tag{1.1}$$

练习 1.1. 证明上述定义的古典概型的概率的确是概率, 即满足定义1.5.

例 1.13. 投掷两枚骰子, 所得数字之和为 7 的概率是多少?

解. 该随机试验的样本空间为 $\Omega = \{(x, y)|x, y = 1, 2, \cdots, 6\}$, 共包含 36 个样本点. 易见, 每个样本点是等可能出现的. 因此这是一个古典概型问题.

所求事件 $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$, 包含 6 个样本点, 因此其概率为

$$P(A) = \frac{6}{36} = \frac{1}{6}.$$ □

例 1.14 (P10)**.** 箱中有 100 件外形一样的同批产品, 其中正品 60 件, 次品 40 件. 现按下列两种方法抽取产品:

(a) **有放回抽样**: 每次任取一件，经观察后放回箱中，再任取下一件;

(b) **无放回抽样**: 每次任取一件，经观察后不放回，在剩下的产品中再任取一件.

试分别用这两种抽样方法，求从这 100 件产品中任意抽取 3 件，其中有两件次品的概率.

解. 易见, 这是古典概型问题. 记事件 A =" 抽得的 3 件产品有两件次品".
(a) 样本空间 Ω_a 包含 100^3 个样本点. 事件 A 中包含的样本点数 $\sharp A = C_3^1 C_{60}^1 40^2$. 因此,

$$P(A) = \frac{C_3^1 C_{60}^1 40^2}{100^3} = 0.288.$$

(b) 样本空间 Ω_b 包含 A_{100}^3 个样本点. 事件 A 中包含的样本点数 $\sharp A = C_3^1 C_{60}^1 A_{40}^2$. 因此,

$$P(A) = \frac{C_3^1 C_{60}^1 A_{40}^2}{A_{100}^3} \approx 0.289.$$ □

一般地, 采用有放回抽样和采用无放回抽样在计算事件的概率时结果是不一样的. 特别地，当被抽取对象的数目较少，差异会更大. 但当被抽取的数目较大，而抽取的数目又较小时，在这两种抽样方式下所计算的概率数值相差不大.

例 1.15 (摸球模型 I)**.** 设箱中有 3 个白球 2 个黑球, 它们除颜色不同外, 其他方面没有区别. 从中任意接连不放回地取出 2 个球, 求第 2 次取出黑球的概率.

错解. 记 W = 白球, B = 黑球, A =" 第 2 次取出黑球". 先后取出 2 个球的样本空间

$$\Omega = \{(W, W), (W, B), (B, W), (B, B)\}.$$

所求事件 $A = \{(W, B), (B, B)\}$. 因此, $P(A) = 2/4 = 1/2$.

解析. 上述解法的错误之处在于这并非古典概型, 因为每个基本事件发生的概率不同:

$$P(W, W) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10}, \quad P(W, B) = \frac{3}{5} \cdot \frac{2}{4} = \frac{3}{10},$$
$$P(B, W) = \frac{2}{5} \cdot \frac{3}{4} = \frac{3}{10}, \quad P(B, B) = \frac{2}{5} \cdot \frac{1}{4} = \frac{1}{10}.$$

事实上,

$$P(A)=P(W,B)+P(B,B)=\frac{2}{5}.$$

然而, 此题仍可归结为古典概型问题: 连续取出 2 个球后, 可继续取出剩余 3 个球, 即将 5 个球连续取出, 求第 2 次刚好取得黑球的概率问题. 如此, 样本空间为

$$\widetilde{\Omega}=\left\{\begin{matrix}(W,W,W,B,B),\ (W,W,B,W,B),\ (W,W,B,B,W),\ (W,B,W,W,B),\ (W,B,W,B,W),\\(W,B,B,W,W),\ (B,B,W,W,W),\ (B,W,B,W,W),\ (B,W,W,B,W),\ (B,W,W,W,B)\end{matrix}\right\},$$

共 10 个样本点. 由于各基本事件发生的概率均为 1/10, 这是个古典概型问题. 又

$$A=\{(W,B,W,W,B),\ (W,B,W,B,W),\ (W,B,B,W,W),\ (B,B,W,W,W)\},$$

因此,

$$P(A)=\frac{4}{10}=\frac{2}{5}.$$ □

上述分析为**列举法**, 适用于情况较为简单的情形. 实际上, 样本空间 $\widetilde{\Omega}$ 中样本点的数量为从 5 个位置中选 3 个位置放置白球的选法数量 C_5^3(剩余位置自动放置黑球). 而事件 A 所包含的样本点数为将 3 白 1 黑共 4 个球置于第 1,3,4,5 个位置 (第 2 个位置留给黑球) 的方法数量 C_4^3, 从而

$$P(A)=\frac{C_4^3}{C_5^3}=\frac{2}{5}.$$

这种分析方法可如下推广至一般情形.

例 1.16 (摸球模型 II)**.** 设箱中有 a 个白球, b 个黑球, 它们除颜色不同外, 其他方面没有区别. 从中任意接连不放回地取出 k 个球 $(k\leqslant a+b)$, 试求第 m $(m\leqslant k)$ 次取出黑球的概率.

解. 记 $A=$" 第 k 次取出黑球", 则仿上例可知

$$P(A)=\frac{C_{a+b-1}^{b-1}}{C_{a+b}^{a}}=\frac{b}{a+b}.$$ □

上例中的结果 $P(A)$ 不依赖于 k 和 m. 由此可知, 彩票抽奖, 抽签等活动中的抽取结果与抽取顺序无关.

概率论中, 许多表现上属性不同的实际问题, 在数学模型上可归结为同一问题. 而解决了一个数学模型后, 可用来解决一类实际问题. 这种应用数学的模型化方法应留心体会.

1.4.2　几何概型

几何概型是指这样的随机试验:

1. 有限性: 试验的结果落于某有限区域 Ω 内;

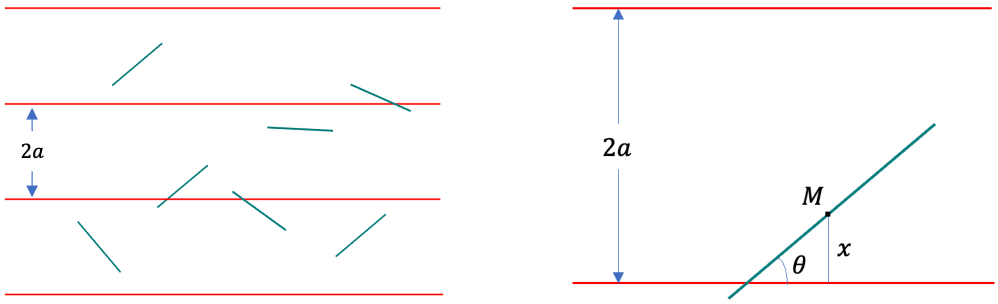

图 1.3: 蒲丰投针问题

2. 等可能性: 试验的结果落于 Ω 中的任何位置都是等可能的.

对任意子区域 $A \subset \Omega$, 试验结果落于 A 的概率为

$$\boxed{P(A) = \frac{\mu(A)}{\mu(\Omega)},}$$

其中, $\mu(A)$ 表示区域 A 的适当测度 (如长度, 面积, 体积等). 实际上, 计数测度

$$\mu(A) = \begin{cases} \sharp A, & A \text{ 包含有限个元素}, \\ \infty, & A \text{ 包含无穷个元素}, \end{cases}$$

是一种特殊的测度. 因此, 几何概型可视作古典概型的推广形式.

例 1.17 (蒲丰投针问题)**.** 在平面上画有等距离的一族平行线 (见图1.3), 平行线间距为 $2a$. 向平面任意投掷一长为 $2l$ $(l < a)$ 的针, 求此针与平行线相交的概率.

解. 此题的关键是将该应用问题数学化. 记该针的中点为 M, 与其最近的直线距离为 x, 夹角为 θ, 见图1.3.

于是, 针的落地姿态可由二维向量 (x, θ) 刻画, 该向量等可能地落于区域

$$\Omega = \{(x, \theta) : 0 \leqslant x \leqslant a,\ 0 \leqslant \theta \leqslant \pi\}.$$

针与 (最近的) 直线相交当且仅当落于子区域

$$A = \{(x, \theta) : 0 \leqslant \theta \leqslant \pi,\ 0 \leqslant x \leqslant l \sin\theta\}.$$

于是, 参考图1.4,

$$P(A) = \frac{A \text{ 的面积}}{\Omega \text{ 的面积}} = \frac{\int_0^\pi l \sin\theta \, \mathrm{d}\theta}{\pi a} = \frac{2l}{\pi a}. \qquad \square$$

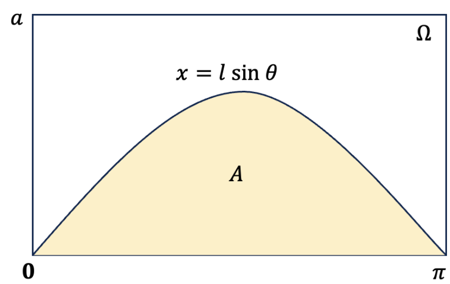

图 1.4: 蒲丰投针几何概型

* 用频率代替概率, 上例的结果可用于估算 π 值:

$$\pi \approx \frac{2l}{a f_N(A)}.$$

例如, Lazzerini (1901) 投针 $N = 34080$ 次, 得估值 $\pi \approx 3.1415929$.