1.1.3 σ 代数*

给定样本空间 Ω, 用 2^{Ω} 表示样本空间 Ω 所有子集所构成的集合, 称为 Ω 的 **幂集**, 即样本空间 Ω 上所有事件所构成的集合. 对可列的样本空间, 将幂集 2^{Ω} 中的元素都看作事件没有任何问题. 但对无限不可列样本空间, 不能简单地将样本空间 Ω 的一切子集都作为事件, 否则将对概率的计算带来困难. 例如, 考虑在区间 $[0,1]$ 上的均匀分布, 任意选取一点, 求所选的点是有理数的概率, 此时难以计算. 为了更好地刻画随机事件, 可以引入可测空间.

定义 1.1 设 Ω 是一个样本空间, 以及 $\Sigma \subseteq 2^{\Omega}$ 是一个由样本空间 Ω 的子集所构成的集合, 若 Σ 满足以下三个条件:

- $\Omega \in \Sigma$;
- 若 $A \in \Sigma$, 则有 $\bar{A} \in \Sigma$;
- 若任意 $A_i \in \Sigma$ $(i = 1, 2, \cdots)$, 则有 $\bigcup_{i=1}^{+\infty} A_i \in \Sigma$,

则称 Σ 是样本空间 Ω 的 σ **代数** (又称 σ **域**), Σ 中的元素称为 **可测集**, 称 (Ω, Σ) 是一个 **可测空间**.

σ 代数 Σ 本质上是一个集合, 其每一个元素也是集合, 即 Σ 是 Ω 一些子集所构成的集合. 若 Σ 是一个 σ 代数, 则 Σ 中每个元素都是可测集. 根据可测空间定义可知

$$\emptyset \in \Sigma, \qquad \bigcup_{i=1}^{n} A_i \in \Sigma, \qquad \bigcap_{i=1}^{n} A_i \in \Sigma, \qquad \bigcap_{i=1}^{+\infty} A_i \in \Sigma.$$

给定样本空间 Ω, 最小的 σ 代数为 $\Sigma = \{\emptyset, \Omega\}$, 最大的 σ 代数为 $\Sigma = 2^{\Omega}$.

若一个非空事件集合 $\mathcal{F} \subset 2^{\Omega}$ 不满足 σ 代数, 则可以构造包含 $\mathcal{F}$ 的最小 σ 代数.

定义 1.2 给定样本空间 Ω 和非空事件集合 $\mathcal{F} \subset 2^{\Omega}$, 记 $\sigma(\mathcal{F})$ 为 **包含 $\mathcal{F}$ 的最小 σ 代数**, 即若 Σ 是一个 σ 代数且 $\mathcal{F} \subset \Sigma$, 则有 $\sigma(\mathcal{F}) \subset \Sigma$.

当 $\mathcal{F} = \{A\}$, 即集合 $\mathcal{F}$ 仅包含单一事件 A, 则最小 σ 代数为 $\sigma(\mathcal{F}) = \{\emptyset, A, \bar{A}, \Omega\}$. 对一般的事件集合 $\mathcal{F}$, 有最小 σ 代数 $\sigma(\mathcal{F}) = \bigcap_{\mathcal{F} \subset \Sigma, \Sigma \text{为} \sigma \text{代数}} \Sigma$.

对有限或可列的样本空间 Ω, 一般考虑 σ 代数 $\Sigma = 2^{\Omega}$; 而当样本空间 Ω 时实数集 $\mathbb{R}$ 时, 一般考虑博雷尔 σ 代数, 即由有限或可列个开区间 (或闭区间) 构成的 σ 代数, 记为 $\mathfrak{R}_1$, 即

$$\begin{aligned}
\mathfrak{R}_1 &= \sigma\left(\{(a,b)\colon a < b \in \mathbb{R}\}\right) = \sigma\left(\{[a,b]\colon a < b \in \mathbb{R}\}\right) = \sigma\left(\{(-\infty, b), b \in \mathbb{R}\}\right) \\
&= \sigma\left(\{(-\infty, b], b \in \mathbb{R}\}\right) = \sigma\left(\{(a, +\infty), a \in \mathbb{R}\}\right) = \sigma\left(\{[a, +\infty), a \in \mathbb{R}\}\right).
\end{aligned}$$

上式中等号成立的原因是 $[a,b] = \bigcap_{n=1}^{\infty}(a + 1/n, b - 1/n)$, $(a,b) = [a,b] \setminus a \setminus b$, 以及可列次的逆、并、交等运算. 类似可定义 n 维博雷尔 σ 代数 $\mathfrak{R}_n$.

1.2 古典概型

古典概型是概率论早期最重要的研究对象, 其发展在概率论中具有重要的意义, 并在产品质量抽样检测等问题中具有广泛的应用. 对古典概型的讨论有助于直观地理解概率论的基本概念, 因而概率的研究通常从古典概型开始.

定义 1.3 (古典概型) 如果试验 E 满足:

- 试验的结果只有有限种可能, 即样本空间 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, 其中 w_i 为基本事件,
- 每种结果发生的可能性相同, 即 $P(\{\omega_i\}) = P(\{\omega_j\})$ $(i \neq j)$,

则称该类试验称为 **古典概型**, 又称 **等可能概型**.

根据上述定义以及 $P(\Omega) = 1$ 可知: 每个基本事件发生的概率为 $P(\{\omega_i\}) = 1/n$, 若事件 A 包含 k 个基本事件 $\{\omega_{i_1}, \omega_{i_1}, \ldots, \omega_{i_k}\}$, 则事件 A 发生的概率为

$$P(A) = k/n = |A|/|\Omega|,$$

这里 $|A|$ 表示事件 A 包含的事件的个数. 根据上式的定义, 不难发现古典概率的一些基本性质.

- **非负性**: 对任意事件 A 有 $P(A) \geqslant 0$;
- **规范性**: $P(\Omega) = 1$;
- **(有限) 可加性**: 若事件 $A_1, A_2, \cdots, A_k$ 两两互不相容, 则有

$$P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k).$$

古典概率计算的本质是计数 (Counting), 计数是组合学研究的重要内容, 后面将在 1.5 节详细介绍各种计数方法, 这里仅介绍一些基本的原理和排列组合.

- **加法原理**: 若一项工作可以用两种不同的过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 完成, 且过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 分别有 n_1 和 n_2 种不同的方法, 则完成该工作有 $n_1 + n_2$ 种不同的方法.
- **乘法原理**: 若一项工作需要依次通过 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 两过程, 且过程 $\mathcal{A}_1$ 和 $\mathcal{A}_2$ 分别有 n_1 和 n_2 种不同的方法, 则完成该工作有 $n_1 \times n_2$ 种不同的方法.

上述两条原理可进一步推广到多个过程的情况.

排列: 从 n 个不同的元素中无放回地取出 r 个元素进行排列, 此时既要考虑取出的元素, 也要顾及排列顺序, 则有 $(n)_r = n(n-1)\cdots(n-r+1)$ 种不同的排列. 若 $r = n$ 时称全排列, 则有 $n!$ 种不同的排列.

组合: 从 n 个不同的元素中无放回地取出 r 个元素, 取出的元素之间无顺序关系, 共有 $\binom{n}{r}$ 种不同的取法, 其中

$$\binom{n}{r} = \frac{n!}{r!(n-r)!} = \frac{(n)_r}{r!}, \quad \text{且记} \quad \binom{n}{0} = 1.$$

这里 $\binom{n}{r}$ 称为 **组合数** 或 **二项系数**, 它是二项展开式 $(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r}$ 中项 $a^r b^{n-r}$ 的系数.

下面介绍一些古典概型的例子:

例 1.4 将 n 个不同的球随机放入 N $(N \geqslant n)$ 个不同的盒子中, 事件 A 表示恰有 n 个盒子中每盒一球; 事件 B 表示指定的 n 个盒子中各有一球; 事件 C 表示指定一盒子恰有 m 个球. 求事件 A, B, C 分别发生的概率. (盒子的容量不限, 放入同一个盒子内的球无顺序排列区别)

解 将 n 个不同的球随机放入 N 个不同的盒子中, 共有 N^n 种不同的放法. 对事件 A, 有 $(N)_n = N!/n!$ 种不同的放法, 因此

$$P(A) = \frac{(N)_n}{N^n} = \frac{N!}{N^n n!}.$$

对事件 B, 有 $n!$ 种不同的放法, 因此

$$P(B) = \frac{n!}{N^n}.$$

对事件 C, 可分为两步: 第一步在指定的盒子内放入 m 个球, 有 $\binom{n}{m}$ 种不同的放法; 第二步将剩下的 $n-m$ 个球放入 $N-1$ 个盒子, 有 $(N-1)^{n-m}$ 种不同的放法. 因此

$$P(C) = \frac{\binom{n}{m}(N-1)^{n-m}}{N^n}.$$

下面介绍古典概型计算中一类典型问题, 在产品质量检测等方面广泛应用.

例 1.5 设一批 N 件产品中有 M 件次品, 现从 N 件产品中不放回地任选 n 件, 求其中恰有 k 件次品的概率.

解 从 N 件产品中任选 n 件, 有 $\binom{N}{n}$ 种不同的选法. 用 A 表示恰有 k 件次品的事件, 即在所选取的 n 件产品中有 k 件次品和 $n-k$ 件正品, 也就是从 M 件次品中选出 k 件次品, 从 $N-M$ 件正品中选出 $n-k$ 件正品, 因此有 $\binom{M}{k}\binom{N-M}{n-k}$ 种不同的取法. 由此可得

$$P(A) = \binom{M}{k}\binom{N-M}{n-k} \Big/ \binom{N}{n}, \tag{1.1}$$

该概率 $P(A)$ 被称为 **超几何概率**.

在上例中采用无放回抽样, 即抽中的产品不再放回, 以后不会被抽到该产品. 下面考虑有放回地抽样, 即被抽中的产品仍放回产品中, 这件产品以后仍然可能被抽到. 例如, 在例 1.5 中若为有放

回地任选 n 件, 则每次抽到一件非次品的概率为 $(N-M)/N$, 抽到一件次品的概率为 M/N, 因此 n 件中恰有 k 件次品的概率为

$$\binom{n}{k}\left(\frac{M}{N}\right)^k\left(\frac{N-M}{N}\right)^{n-k}.$$

抽签是人们引入随机性的一个通用例子, 广泛应用于各种体育赛事或日常生活中. 关于抽签的公平性, 即抽签结果虽然不同但出现这种结果的可能性相同, 需要通过计算概率来进行验证:

例 1.6 (抽签问题) 盒子中有 a 个不同的白球, b 个不同的红球, 假设有 k 个人依次随机无放回地从盒中取一个球, 问第 i 个人 $(i \leqslant k)$ 取出红球的概率是多少?

解 若 k 个人依次随机无放回地从袋中取一个球, 则有 $(a+b)_k$ 种不同的取法. 用 A 表示第 i 个人取到红球的事件, 若 A 发生, 第 i 个人取到红球, 它可能是 b 个红球中的任意一个, 有 b 种取法; 其它剩余的 $k-1$ 个人可以从 $a+b-1$ 个球中随机取出, 有 $(a+b-1)_{k-1}$ 种不同的取法. 因此事件 A 发生的概率为

$$P(A)=\frac{b(a+b-1)_{k-1}}{(a+b)_k}=\frac{b}{a+b}.$$

由此可知第 i 个人取到红球的概率为 $b/(a+b)$, 与 i 的大小无关, 即抽签先后顺序对抽签的结果没有影响, 由此证明了抽签的公平性.

根据古典概率的规范性和可加性, 对任意事件 A 有 $1=P(\Omega)=P(A\cup\bar{A})=P(A)+P(\bar{A})$ 成立, 由此可得

$$P(\bar{A})=1-P(A) \quad 和 \quad P(\emptyset)=0.$$

灵活运用这些性质, 可以有效简化古典概率的计算. 例如,

例 1.7 (De Mere问题) 用 A 表示一枚骰子连续任意投掷 4 次得到一点的事件, 用 B 表示两个骰子连续任意投掷 24 次得到两个一点的事件, 求事件 A 和 B 哪个发生的概率大?

解 根据题意可知对立事件 $\bar{A}$ 表示在 4 次投掷中都没有出现一点, 因此不难得到事件 A 不发生的概率 $P(\bar{A})=(5/6)^4$, 根据概率性质有

$$P(A)=1-P(\bar{A})=1-(5/6)^4=0.5177.$$

同理可得对立事件 $\bar{B}$ 发生的概率为 $P(\bar{B})=(35/36)^{24}$, 于是有

$$P(B)=1-P(\bar{B})=1-(35/36)^{24}=0.4914.$$

因此事件 A 发生的概率更大些.

该问题是 17 世纪中期 De Mere 向帕斯卡 (Pascal) 请教的著名问题之一. 为了求解这些问题, 帕斯卡与费马 (Fermat) 进行了信件交流, 由此提出了概率的若干原理, 奠定了概率论的基础.

生日问题是概率历史上有名的数学问题, 研究某次集会的 k 个人中至少有两人生日相同的概率, 或一个包含 k 个人的班级中至少两人生日相同的概率.

例 1.8 (生日问题) 有 k 个人 ($k < 365$), 每个人的生日等可能地出现于 365 天中的任意一天, 求至少两人生日相同的概率.

解 用 A 表示至少有两人生日相同的事件, 其对立事件 $\bar{A}$ 表示任意两人生日均不相同的事件. k 个人的生日共有 365^k 种可能, 而 k 个人的生日两两互不相同的有 $(365)_k$ 种可能. 因此

$$P(A) = 1 - P(\bar{A}) = 1 - \frac{(365)_k}{365^k}.$$

根据该公式给出部分 k 和 $P(A)$ 的如下数值结果:

k	10	20	23	30	40	50	60	64	100
$P(A)$	0.117	0.411	0.507	0.706	0.891	0.97	0.994	0.997	0.9999997

由此表可知, 当一个班级超过 23 人时, 同生日的概率超过 1/2; 当超过 60 人时, 则几乎能肯定至少有二人同生日.

1.3 几何概型

古典概型考虑有限的样本空间, 即只有有限个基本事件, 然而在很多实际应用中可能遇见具有无限多基本事件的等可能性现象, 具有如下两个特点:

- **样本空间无限可测** 样本空间包含无限不可列个样本点, 但可以用几何图形 (如一维线段、二位平面区域、或三维空间区域等) 来表示, 其相应的几何测度 (如长度、面积、体积等) 是一个非零有限的实数,
- **基本事件等可能性** 每个基本事件发生的可能性大小相等, 从而使得每个事件发生的概率与该事件的几何测度相关, 与具体位置无关,

则称这类试验为 **几何概型**. 形式化定义如下:

定义 1.4 在测度有限的区域 Ω 内等可能性投点, 落入 Ω 内的任意子区域 A 的可能性与 A 的测度成正比, 与 A 的位置与形状无关, 这样的概率模型称之为**几何概型**, 事件 A 发生的概率为

$$P(A) = \frac{A\text{ 的测度}}{\Omega\text{ 的测度}} = \frac{\mu(A)}{\mu(\Omega)}.$$

根据上述定义, 不难发现几何概率的一些基本性质.

- **非负性**: 对任意事件 A 有 $P(A) \geqslant 0$;

- **规范性**: $P(\Omega) = 1$;
- **可列可加性**: 若可列个事件 $A_1, A_2, \cdots, A_n, \cdots$ 两两互不相容, 则有

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

这里的可列可加性是根据测度的可列可加性 $\mu(\cup_{i=1}^{+\infty} A_i) = \sum_{i=1}^{+\infty} \mu(A_i)$. 下面给出几何概型的例子.

例 1.9 假设乘客到达汽车站的时间是等可能的, 客车间隔一段时间发班, 请规划最长的间隔发车时间, 才能确保乘客候车等待时间不超过 20 分钟的概率大于 80%.

解 设客车的间隔时间为 l $(l > 20)$, 则在一个间隔为 l 分钟的时间内, 所有乘客是等可能到达的, 因此考虑乘客到达时间的样本空间 $\Omega = \{x\colon 0 < x \leqslant l\}$. 用 B 表示乘客的等待时间超过 20 分钟的事件, 而事件 B 发生则可知乘客到达车站的时间在 0 与 $l - 20$ 之间, 即

$$B = \{x\colon 0 < x < l - 20\}.$$

可知事件 B 发生的概率小于或等于 20%, 即

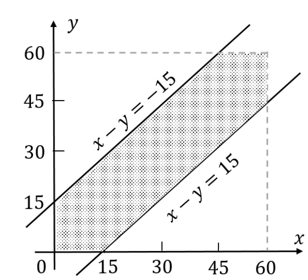
$$P(B) = \frac{l - 20}{l} \leqslant 0.2,$$

求解可得 $l \leqslant 25$.

例 1.10 (会面问题) 两人约定中午 $12:00 - 13:00$ 到某地会面, 两人在这时间内到达的时间是等可能的, 先到者等另一人 15 分钟后离开, 求两人见面的概率.

解 首先用 x, y 分别表示两人的到达时间, 则整个样本空间 $\Omega = \{(x, y)\colon 0 \leqslant x, y \leqslant 60\}$. 用 A 表示两人见面的事件, 则

$$A = \{(x, y)\colon |x - y| \leqslant 15\} = \{(x, y)\colon x - y \leqslant 15 \text{ 且 } x - y \geqslant -15\}.$$



如右图所示, 阴影部分表示事件 A 发生的区域, 根据几何概型有

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{60^2 - 45^2}{60^2} = \frac{7}{16}.$$

思考: 若两银行经理非常聪明且希望能促成此次见面, 但没有通讯方式进行联系, 能否找出一些策略来解决会面问题?

很多几何概型的概率可通过计算机模拟仿真来近似计算, 称为 **统计模拟法** 或 **蒙特卡洛 (Monte Carlo) 法**: 先构造相应的概率模型, 再利用计算机模拟试验, 用统计的方法计算其估计值, 作为所求问题的近似值. 例如, 可利用蒙特卡洛法来近似计算例 1.10 的概率, 伪代码为:

输入参数: 试验总次数 N.	%% 取较大正整数 N, 能更精确计算概率
初始化: 事件 A 最初始发生的次数 $n_A \leftarrow 0$.	%% 此处事件 A 表示两人见面的事件
For $i = 1 : N$	
$x \leftarrow \mathrm{Random}(0, 60)$, $y \leftarrow \mathrm{Random}(0, 60)$.	%% 在区间 $(0, 60)$ 以随机任意选取两个数
If $\lvert x - y\rvert \leqslant 15$ then	
$n_A \leftarrow n_A + 1$.	%% 若两人见面则频数+1
End	
End	
输出概率: n_A/N.	

接下来介绍几何概型的一个经典问题, 由法国科学家蒲丰于 1777 年提出.

例 1.11 (投针问题) 平面上有两条平行线, 相距为 a, 向此平面任投一长度为 l $(l < a)$ 的针, 求此针与任一平行线相交的概率.

解 用 x 表示针的中点到最近一条平行线的距离, 以及用 θ 表示针与平行线的夹角, 针与平行线的位置关系如图 1.2(a) 所示. 很容易知道 $x \in [0, a/2]$ 和 $\theta \in [0, \pi]$, 因此整个样本空间 $\Omega = \{(x, \theta) \colon x \in [0, a/2], \theta \in [0, \pi]\}$.

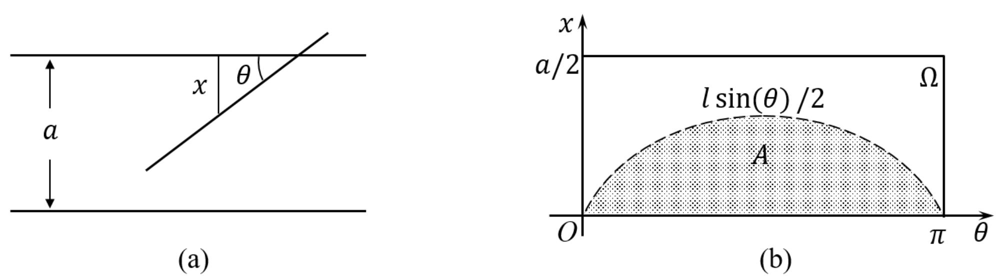


图 1.2 蒲丰投针问题

用 A 表示针与平行线相交的事件, 若事件 A 发生则必有 $x \leqslant l\sin(\theta)/2$ 成立. 如图 1.2(b) 所示, 阴影部分表示事件 A 发生的区域, 根据几何概型有

$$P(A) = \frac{\mu(A)}{\mu(\Omega)} = \frac{A \text{ 的面积}}{\Omega \text{ 的面积}} = \frac{\int_0^\pi l\sin(\theta)/2d\theta}{a\pi/2} = \frac{2l}{a\pi}\ . \tag{1.2}$$

在上例中, 事件 A 发生的概率与圆周率 π 有关, 由此蒲丰设想出计算 π 的概率近似方法, 通过频率来近似计算事件 A 发生的概率, 再根据 (1.2) 计算圆周率 π, 即

$$\frac{n_A}{n} \approx P(A) = \frac{2l}{a\pi} \qquad \Rightarrow \qquad \pi \approx \frac{2ln}{an_A}\ .$$

这里 n 表示试验的总次数, 而 n_A 表示事件 A 发生的次数. 历史上有多人根据蒲丰的设想做了试验来近似计算圆周率 π, 这里给出了部分的实验结果.

实验者	年份	针长	投掷次数	相交次数	π 的估计值
Wolf	1850	0.8	5000	2532	3.1596
Smith	1855	0.6	3204	1218.5	3.1554
Fox	1884	0.75	1030	489	3.1595
Lazzerini	1901	0.83	3408	1808	3.1415929
Reina	1925	0.5419	2520	859	3.1795

尽管看起来不是一种估计圆周率 π 的好方法, 即使在当时用其它数学方法也可以准确估计到圆周率 π 小数点后面一百位了. 但是提出了一种新的估计方法: 通过建立概率模型, 设计合适的随机实验, 通过统计模拟来确定计圆周率的值, 蒲丰投针开创了统计模拟法的先河.

在二十世纪之前, 很多人都相信只要找到合适的等可能性描述, 概率是可以被唯一定义的. 然而贝特朗 (Joseph Bertrand, 1822-1900) 在 1889 年对这种观点提出了质疑, 他通过下面的一个具体例子说明: 几何概型的等可能性概率存在多种看似合理但相互矛盾的结果.

例 1.12 (贝特朗奇论) 在半径为 1 的圆内随机地取一条弦, 求其弦长超过该圆内接等边三角形边长 $\sqrt{3}$ 的概率.

解 对于"等可能性"或"随机性"含义的不同解释, 这个问题存在着多种不同答案的解决方法, 下面给出三种不同的方法:

(1) 任何与圆相交的弦一般有两个交点, 不妨在圆上先固定其中一点, 以此点为顶点作一个等边三角形, 只有落入此三角形内的弦才满足弦长超过 $\sqrt{3}$. 这种弦的另一端跑过的弧长为整个圆周的 1/3, 故所求概率等于 1/3 (如图 1.3-**a**).

(2) 弦长只跟到圆心的距离有关, 与方向无关, 因此可以假定它垂直于某一条直径. 当且仅当它与圆心的距离小于 1/2 时, 其弦长才会大于 1/3, 因此此时所求的概率为 1/2 (如图 1.3-**b**).

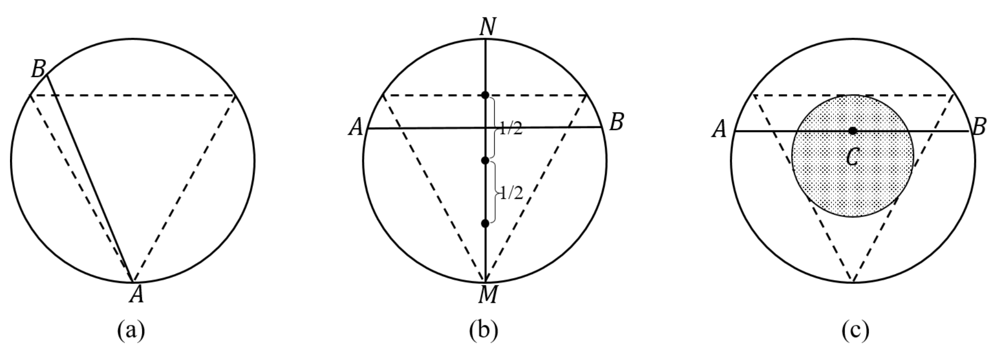


图 1.3 贝特朗奇论