

概率论与数理统计



第

2

讲

古典概率、几何概率、频率

在日常生活中，大量重复一随机试验时，会发现，有些事件发生的次数多一些，有些事件发生的次数少一些。

导致这一现象的原因是事件发生的可能性有大小之分。

将表征随机事件发生可能性大小的数称为事件的概率

记为 $P(A)$



如何度量概率的值呢?

假定随机试验 E 有有限个可能的结果 $\omega_1, \omega_2, \dots, \omega_n$

假定从该试验的条件及实施方法上去分析，找不到任何理由认为其中某一结果出现的机会比另一结果出现的机会大或小，只好认为所有结果在试验中有同等可能的出现机会，即 $1/n$ 的出现机会。

满足这两个条件的随机试验称为**古典概型**。

若随机试验 E 满足：

- (1) 试验的样本空间 S 只含有有限个元素 $S = \{\omega_1, \dots, \omega_n\}$;
- (2) 试验中每个基本事件的发生是等可能的。

则称 E 为等可能概型，简称古典概型。

设随机试验 E 的样本空间 S 含有 n 个样本点，事件 A 包含 k 个样本点，定义

$$P(A) = \frac{k}{n}$$

称此概率为古典定义。

计算方法:

- (1) 当样本点的总数较小时, 列出 S 和 A 中的样本点, 然后数点的个数。
- (2) 当样本点的总数较大时, 需要以排列组合为工具计算 S 和 A 中的样本点的个数。

例1. 将一枚硬币抛掷两次, 求出现一个正面和一个反面的概率。

解: 易知此试验的样本空间为 $S=\{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{正}), (\text{反}, \text{反})\}$

设 A 表示出现一个正面一个反面。则 $A=\{(\text{正}, \text{反}), (\text{反}, \text{正})\}$ 。

所以 $P(A)=\frac{2}{4}=\frac{1}{2}$

若 $S=\{(\text{正}, \text{正}), (\text{正}, \text{反}), (\text{反}, \text{反})\}$ 是否合适?

基本计数原理

1. 加法原理:

设完成一件事有 m 种方式, 第一种方式有 n_1 种方法,
第二种方式有 n_2 种方法, \dots ; 第 m 种方式有 n_m 种方法,
无论通过哪种方法都可以完成这件事。

则完成这件事总共有 $n_1 + n_2 + \dots + n_m$ 种方法。

基本计数原理

2. 乘法原理:

设完成一件事有 m 个步骤, 第一个步骤有 n_1 种方法,
第二个步骤有 n_2 种方法, ...; 第 m 个步骤有 n_m 种方法,
必须通过每一步骤, 才算完成这件事。

则完成这件事总共有 $n_1 \times n_2 \times \dots \times n_m$ 种方法。

从 n 个不同元素(无放回地)取 k 个元素($1 \leq k \leq n$)排成一列, 方法总数为

$$A_n^k = n \cdot (n-1) \cdots (n-k+1)$$

从 n 个不同元素有放回地依次取 k 个元素排成一列, 方法总数为

$$n \cdot n \cdots n = n^k$$

从 n 个不同的元素中(无放回地)抽取 k 个元素不论次序地组成一组, 方法总数为

$$C_n^k = \frac{n!}{k!(n-k)!}$$

例1. 一袋中有10个相同的球，分别标有号码1, 2, ..., 10。今不放回抽取3个球，求这3个球的号码均为偶数的概率。

解一： 设 $A = \{ \text{不放回抽取的3个球的号码均为偶数} \}$

试验为从袋中不放回抽取3个球，这相当于从10个不同的数中任取3个作排列，排列数为 $n = A_{10}^3 = 10 \times 9 \times 8$

事件 A 相当于从5个偶数中取3个排列，排列数为 $k = A_5^3 = 5 \times 4 \times 3$

$$P(A) = \frac{A_5^3}{A_{10}^3} = \frac{1}{12}$$

例1. 一袋中有10个相同的球，分别标有号码1, 2, ..., 10。今不放回抽取3个球，求这3个球的号码均为偶数的概率。

解二：“不放回抽取3个球”可以理解为，“从袋中一次性取出3个球”。

由此，从袋中不放回抽取3个球，相当于从10个不同的数中任取3个组合成1组，组合数为 $n = C_{10}^3$

“一把抓”问题

事件A相当于从5个偶数中取3个组合成1组，组合数为 $k = C_5^3$

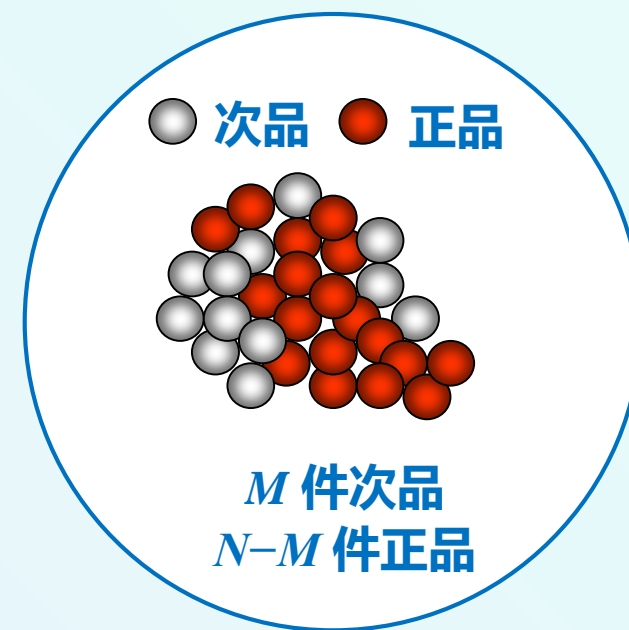
$$P(A) = \frac{C_5^3}{C_{10}^3} = \frac{1}{12}$$

例2. 设一批同类型的产品共有 N 件，其中次品有 M 件。今从中任取 n （假定 $n \leq N - M$ ）件，求次品恰有 k 件的概率 ($0 \leq k \leq \min(M, n)$)。

解：令 $B = \{\text{恰有 } k \text{ 件次品}\}$ ，取出 n 件产品的方法总数为 C_N^n

对于事件 B ：第一步，先从 M 中取出 k 件次品，第二步，从 $N - M$ 中取出 $n - k$ 件正品，方法总数为 $C_M^k C_{N-M}^{n-k}$

$$\therefore P(B) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$$



古典概率的性质

设 E 是古典概型, 其样本空间为 $S = \{\omega_1, \dots, \omega_n\}$; A, A_1, A_2, \dots, A_m 是 E 中的事件, 则有

(1) 非负性: $P(A) \geq 0$;

(2) 归一性: $P(S) = 1$;

(3) 可加性: 若 A_1, A_2, \dots, A_m 是互不相容的事件, 则有 $P(\bigcup_{i=1}^m A_i) = \sum_{i=1}^m P(A_i)$

推论: $P(A) = 1 - P(\bar{A})$

例4.从10件同类型产品(其中6件正品, 4件次品)中任取4件, 求(1) “取得的4件中次品不超过1件” (事件 A)的概率, (2) “取得的4件中次品至少有1件” (事件 B)的概率。

解: 设 A_i 为 “取得的4件中恰好有 i 件次品”, $i=0,1,2,3,4$ 。则有

(1) $A=A_0 \cup A_1$ 且 A_0, A_1 互不相容

$$P(A)=P(A_0 \cup A_1)=P(A_0)+P(A_1)=\frac{C_6^4}{C_{10}^4}+\frac{C_6^3 \times C_4^1}{C_{10}^4}=\frac{19}{42}$$

例4.从10件同类型产品(其中6件正品, 4件次品)中任取4件, 求(1) “取得的4件中次品不超过1件” (事件 A)的概率, (2) “取得的4件中次品至少有1件” (事件 B)的概率。

解: 设 A_i 为 “取得的4件中恰好有 i 件次品”, $i=0,1,2,3,4$ 。则有

(2) $B=A_1 \cup A_2 \cup A_3 \cup A_4$ 且 A_1, A_2, A_3, A_4 互不相容

$$P(B)=P(A_1 \cup A_2 \cup A_3 \cup A_4)=P(A_1)+P(A_2)+P(A_3)+P(A_4)=\sum_{i=1}^4 \frac{C_6^{4-i} \times C_4^i}{C_{10}^4} = \frac{13}{14}$$

$$P(B)=1-P(\bar{B})=1-P(A_0)=1-\frac{C_6^4}{C_{10}^4}=\frac{13}{14}$$

例5. 设有 n 个球，每个球都以同样的概率 $1/N$ 落入到 N 个格子($N \geq n$)的任意一个格子，试求 (1) n 个球放到任意 n 个格子中且各有一球(A)的概率； (2) 某指定的一个盒子中没有球(B)的概率。

解： (1) 将 n 个球投入 N 个格中，共有 N^n 种放法

A ： 第一步，先选出 n 个格子，共有 C_N^n 种选法

第二步，对于选中的 n 个格子，各放一个球，共有 $n!$ 种放法。

即： A 中有 $C_N^n n!$ 个样本点。 所以 $P(A) = \frac{C_N^n n!}{N^n}$

例5. 设有 n 个球，每个球都以同样的概率 $1/N$ 落入到 N 个格子($N \geq n$)的任意一个格子，试求 (1) n 个球放到任意 n 个格子中且各有一球(A)的概率； (2) 某指定的一个盒子中没有球(B)的概率。

解： (2) 将 n 个球投入 N 个格中，共有 N^n 种放法

B ：指定的格子中不能放球，因此， n 个球中的每一个球可以并且只可以放入其余的 $N-1$ 个格子中。总共有 $(N-1)^n$ 种放法。

$$\text{所以 } P(B) = \frac{(N-1)^n}{N^n}$$

生日问题 某人群有 $n(n < 365)$ 个人，求任何两个人生日都不相同(A)的概率。

解：试验相当于从365天中可重复地找 n 天，分别作为 n 个人的生日，它相当于 n 个球扔入365个格子中。

任何两个人生日都不相同，相当于把 n 个球放入到 n 个格子中，且每个格子各有一个球。

$$\text{所以 } P(A) = \frac{C_{365}^n n!}{365^n}$$

$$\text{至少有两个人生日相同的概率为: } P(\bar{A}) = 1 - P(A) = 1 - \frac{C_{365}^n n!}{365^n}$$

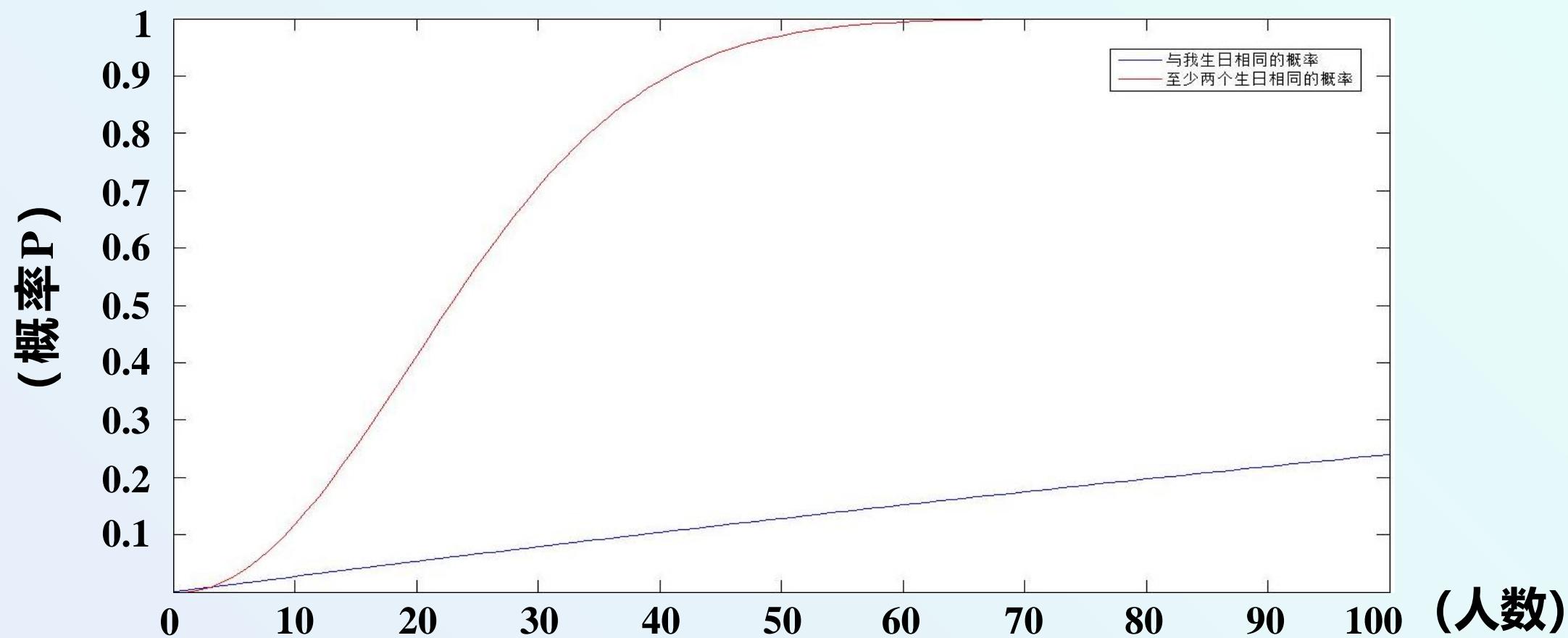
生日悖论 某人群有 n 个人, 求他们中有人与“我”生日相同(B)的概率?

解: B 的对立事件为无人与“我”生日相同

相当于某指定的一个格子中没有球

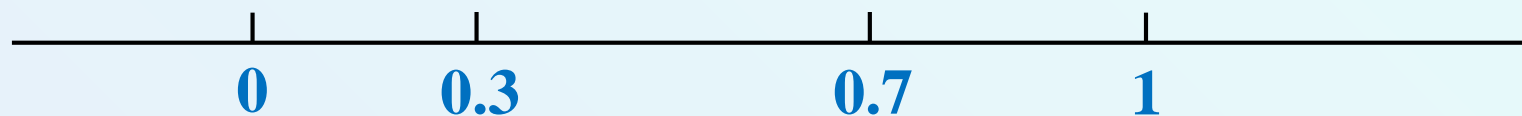
$$\text{所以: } P(\bar{B}) = \frac{(365-1)^n}{365^n} \quad \text{即有: } P(B) = 1 - \frac{(365-1)^n}{365^n}$$

人数	至少有两同生日相同的概率	有人与“我”生日相同的概率
10	0.116948	0.027062
20	0.411438	0.053391
30	0.706316	0.079009
40	0.891231	0.103932
50	0.970474	0.128282
60	0.994122	0.151774
70	0.999159	0.174729
80	0.999914	0.197063
90	0.999993	0.218791
100	0.999999	0.239932



在概率论发展早期，人们就已经注意到只考虑那种仅有有限个样本点的随机试验是不够的，还必须考虑试验结果是无穷多个的情形，这中间最简单的一类是试验结果有**无穷多个**，而又有某种 **“等可能性”** 的情形。

如：向 $[0,1]$ 区间内随机的投点。



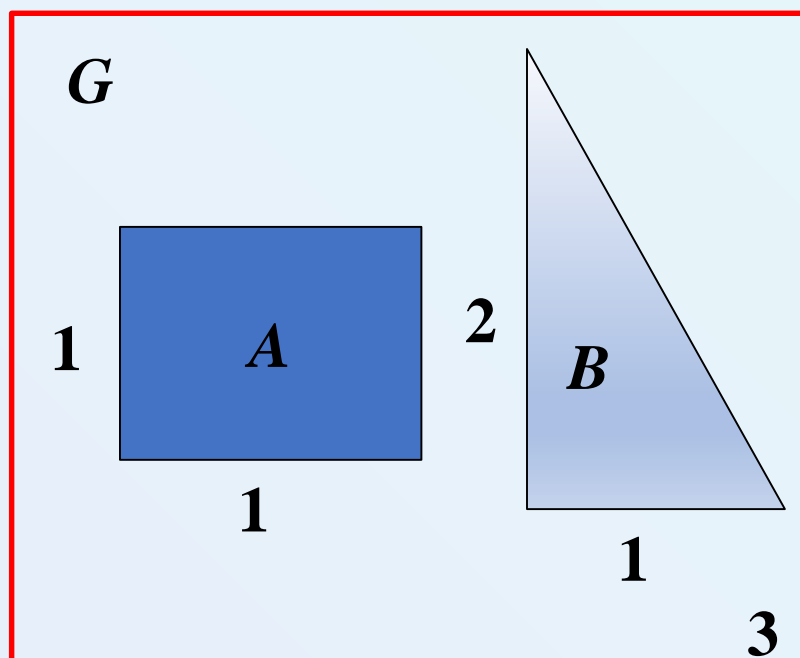
问：事件 A ，点落在0与0.3之间的概率。

$$P(A) = \frac{\text{事件}A\text{对应的区间长度}}{\text{整个区间的长度}} = 0.3$$

事件 B ，点落在0.7与1之间的概率。

$$P(B) = \frac{\text{事件}B\text{对应的区间长度}}{\text{整个区间的长度}} = 0.3$$

在如，向如下正方形区域 G 内投点，求点落在区域 A 和 B 内的概率 $P(A)$ 和 $P(B)$.



$$P(A) = \frac{\text{区域}A\text{的面积}}{\text{区域}G\text{的面积}} = \frac{1}{9}$$

$$P(B) = \frac{\text{区域}B\text{的面积}}{\text{区域}G\text{的面积}} = \frac{1}{9}$$

这种与几何测量有关的概率称为**几何概率**.

1. 几何概型定义

向任一可度量区域 G 内投一点，如果所投的点落在 G 中任意可度量区域 g 内的可能性与 g 的度量成正比，而与 g 的位置和形状无关，则称这个随机试验为几何概型的随机试验，或简称为几何概型。

2. 几何概率的定义

在几何概型中，样本空间为 $S=G$ ， G 中的点是样本点

设 $A=\{\text{投点落入区域}g\text{内}\}$ ，则有 $P(A)=[g\text{的度量}] / [G\text{的度量}]$

几何概率的“等可能性”可解释为“等度量（长度、面积、体积）等概率”

例6. 两人约定于12点到1点到某地会面，先到者等20分钟后离去，试求两人能会面的概率？

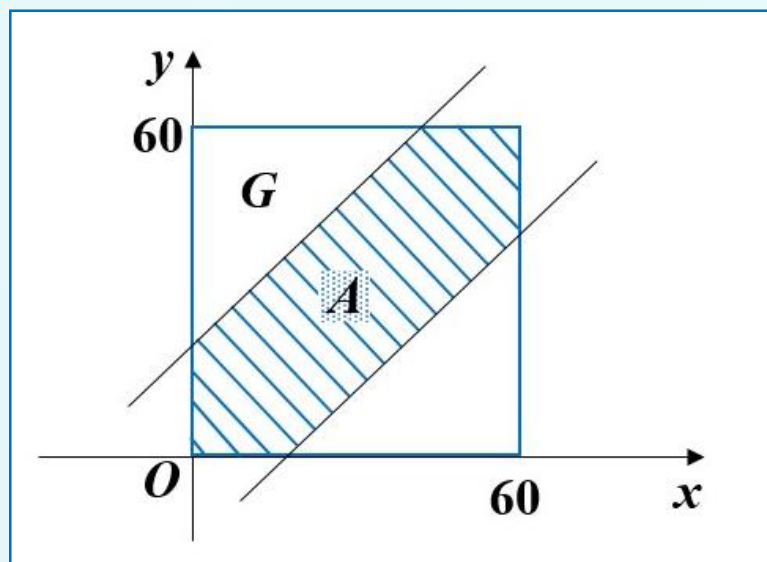
解：设 x, y 分别为两人到达的时刻

甲乙二人分别“等可能”地在12点—1点的任何时刻达到，故可看作几何概型。

设 A 表示“两人能会面”，则有

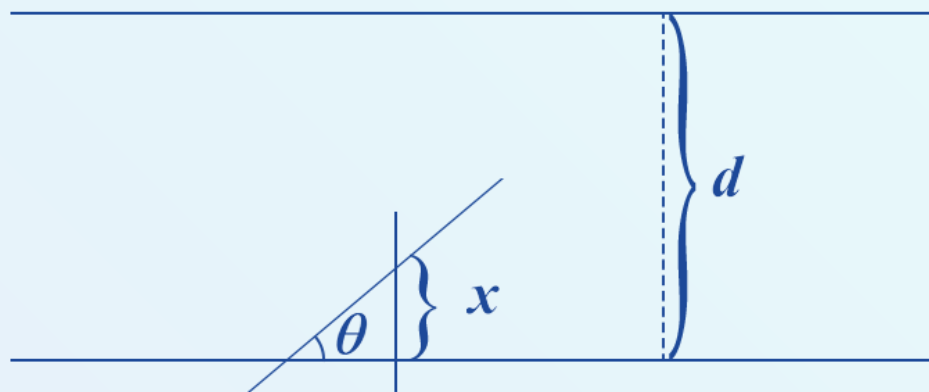
$$A = \{ (x, y) : |x - y| \leq 20 \}$$

$$\text{所以 } P(A) = \frac{60^2 - 40^2}{60^2} = \frac{5}{9}$$



例7. (蒲丰投针问题) 设平面上画满了间距为 d 的等距平行线, 向该平面内任意投掷一枚长度为 l ($l < d$) 的针, 求该针与任一平行线相交的概率。

解: 设 x 表示针的中点与最近一条平行线的距离, θ 表示针与此线的夹角

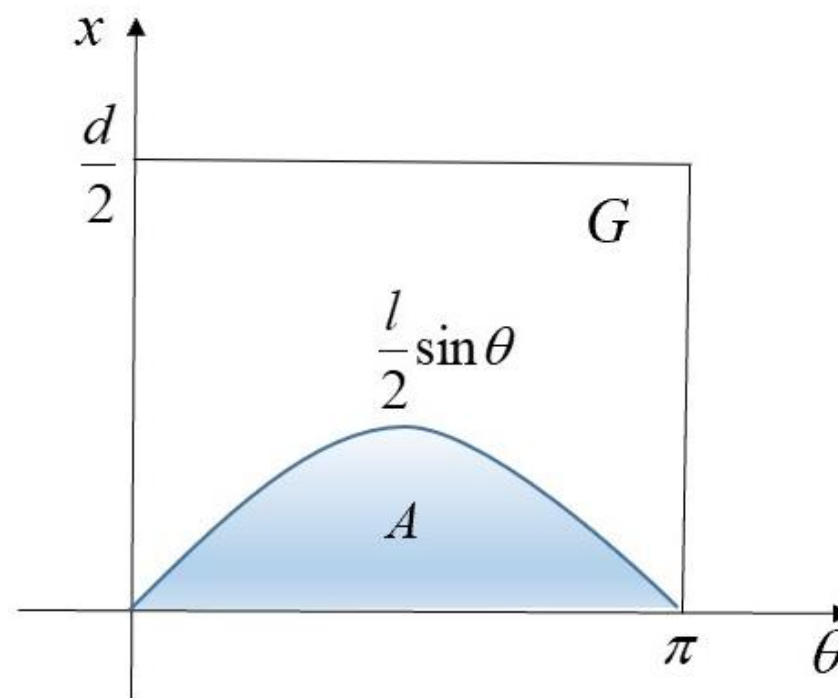


则样本空间为: $G = \{(x, \theta) : 0 < x < \frac{d}{2}, 0 < \theta < \pi\}$ 且 $S_G = \frac{d\pi}{2}$

设事件 A 表示针与一条平行线相交。易知满足的条件为

$$x < \frac{l}{2} \sin \theta \quad \text{且} \quad S_A = \int_0^{\pi} \frac{l}{2} \sin \theta d\theta = l$$

所以
$$P(A) = \frac{S_A}{S_G} = \frac{l}{d\pi/2} = \frac{2l}{d\pi}$$



几何概型的基本性质

设 A, A_1, A_2, \dots, A_n 是 E 中事件, 则有

(1) 非负性: $P(A) \geq 0$;

(2) 归一性: $P(S) = 1$;

(3) 可加性: 若 A_1, A_2, \dots, A_n 是互不相容的事件, 则有 $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$

在古典概率和几何概率的定义中，“等可能性”是一基本假定。
然而在许多实际问题中，这一基本假定并不成立。

例如：从同一型号的反坦克弹中任取一发射击目标，观察命中情况。设 $A=\{\text{命中}\}$ 。求 $P(A)$ 。

样本空间为 $S=\{\text{命中}, \text{不命中}\}$

此试验既不是古典概型，也不是几何概型。

设随机试验 E 的样本空间为 S , A 为 E 的一个事件。在相同的条件下, 将试验 E 重复进行 n 次, 在这 n 次试验中, 事件 A 发生的次数 n_A 称为事件 A 发生的**频数**。比值 n_A/n 称为在这 n 次试验中, 事件 A 发生的**频率**, 记作 $f_n(A)$ 。

$$f_n(A) = \frac{n_A}{n} \rightarrow$$

表示事件 A 发生的频繁程度

直观想法: 用事件 A 频率来近似概率, 但是否可行呢?

频率稳定性

例8. 掷一枚有正反两面的均匀硬币，观测正面朝上的次数并计算其频率。将一枚硬币分别抛掷5次，20次，50次，100次，500次为一组，每组各做20次试验，试验结果见下表，其中 n_A 表示正面朝上的频数， $f(A)$ 表示正面朝上的频率。



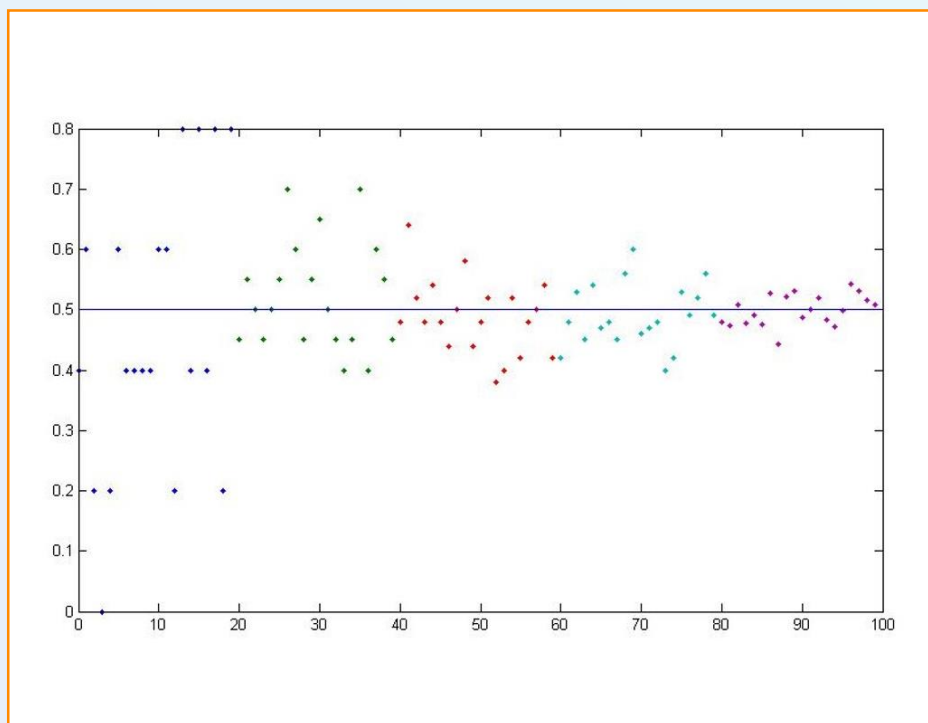
序号	n=5		n=20		n=50		n=100		n=500	
	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$
1	2	0.4	9	0.45	24	0.48	42	0.42	240	0.480
2	3	0.6	11	0.55	32	0.64	48	0.48	237	0.474
3	1	0.2	10	0.50	26	0.52	53	0.53	254	0.508
4	0	0	9	0.45	24	0.48	45	0.45	239	0.478
5	1	0.2	10	0.50	27	0.54	54	0.54	245	0.490
6	3	0.6	11	0.55	24	0.48	47	0.47	238	0.476
7	2	0.4	14	0.70	22	0.44	48	0.48	264	0.528
8	2	0.4	12	0.60	25	0.50	45	0.45	222	0.444
9	2	0.4	9	0.45	29	0.58	57	0.56	261	0.522
10	2	0.4	11	0.55	22	0.44	60	0.60	266	0.532

三. 频率

序号	$n=5$		$n=20$		$n=50$		$n=100$		$n=500$	
	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$	n_A	$f(A)$
11	3	0.6	13	0.65	24	0.48	46	0.46	244	0.488
12	3	0.6	10	0.50	26	0.52	47	0.47	250	0.500
13	1	0.2	9	0.45	19	0.38	48	0.48	260	0.520
14	4	0.8	8	0.40	20	0.40	40	0.40	242	0.484
15	2	0.4	9	0.45	26	0.52	42	0.42	236	0.472
16	4	0.8	14	0.70	21	0.42	53	0.53	249	0.498
17	2	0.4	8	0.40	24	0.48	49	0.49	271	0.542
18	4	0.8	12	0.60	25	0.50	52	0.52	266	0.532
19	1	0.2	11	0.55	27	0.54	57	0.56	258	0.516
20	4	0.8	9	0.45	21	0.42	49	0.49	254	0.508

频率稳定性

所有的值都以0.5为中心上下波动。



$n=5$ ，频率的最小值为0，最大值为0.8，波动的幅度较大；

$n=20$ ，最小值为0.4，最大值为0.7，波动的幅度变小了；

$n=50$ ，最小值为0.38，最大值为0.64，波动的幅度进一步变小；

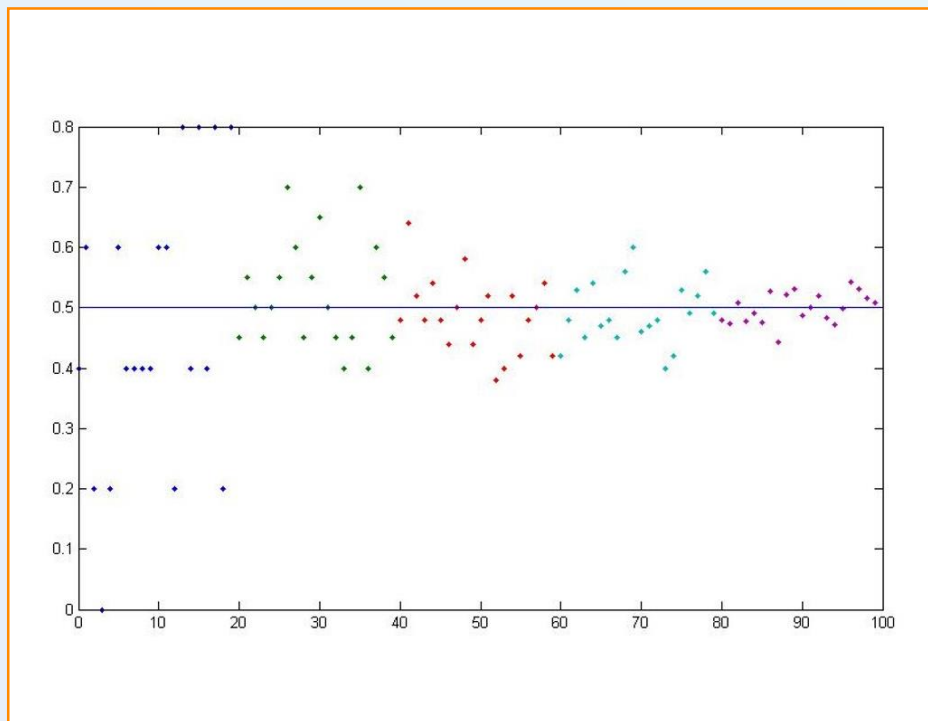
$n=100$ ，最小值为0.40，最大值为0.60，波动的幅度更小了；

频率稳定性

所有的值都以0.5为中心上下波动。

当 n 增大到500时，最小值和最大值为0.444和0.542。

波动的幅度已经非常小了，也就是说随着 n 的增大，频率的波动幅度越来越小，逐渐稳定于0.5。



频率稳定性



长期实践表明，在相同的条件下，重复进行试验，事件 A 发生的频率 $f_n(A)$ 总在一个常数值附近波动，而且随着重复试验次数 n 的增加，频率的波动幅度越来越小。观测到的大偏差越来越少，呈现出一定的稳定性。这就是**频率的稳定性**的特点。

频率稳定性

实验者	投币次数 n	正面次数 n_A	频率 f_A
德·摩根 (DeMorgan)	2048	1061	0.5181
蒲丰 (Buffon)	4040	2048	0.5069
克里奇 (J.Kerrich)	7000	3516	0.5022
克里奇 (J.Kerrich)	8000	4034	0.5042
克里奇 (J.Kerrich)	9000	4538	0.5042
克里奇 (J.Kerrich)	10000	5067	0.5067
费勒 (Feller)	10000	4979	0.4979
卡尔·皮尔逊K.Pearon	12000	6019	0.5016
卡尔·皮尔逊K.Pearon	24000	12012	0.5005
罗曼诺夫斯基	80640	40173	0.4982

频率的基本性质

设 A, A_1, A_2, \dots, A_m 是 E 中事件, 则有

(1) 非负性: $f_n(A) \geq 0$;

(2) 归一性: $f_n(S) = 1$;

(3) 可加性: 若 A_1, A_2, \dots, A_m 是互不相容的事件, 则有 $f_n(\bigcup_{i=1}^m A_i) = \sum_{i=1}^m f_n(A_i)$

概率的频率定义

在一组不变的条件下，重复作 n 次试验，记 n_A 是 n 次试验中事件 A 发生的次数。当试验次数 n 很大时，频率 $f_n(A)=n_A/n$ 稳定地在某数值 p 附近波动，而且一般地说，随着试验次数的增加，这种波动的幅度越来越小，称数值 p 为事件 A 在这一组不变的条件下发生的概率，记作 $P(A)=p$ 。也称这个概率为统计概率。

意义：

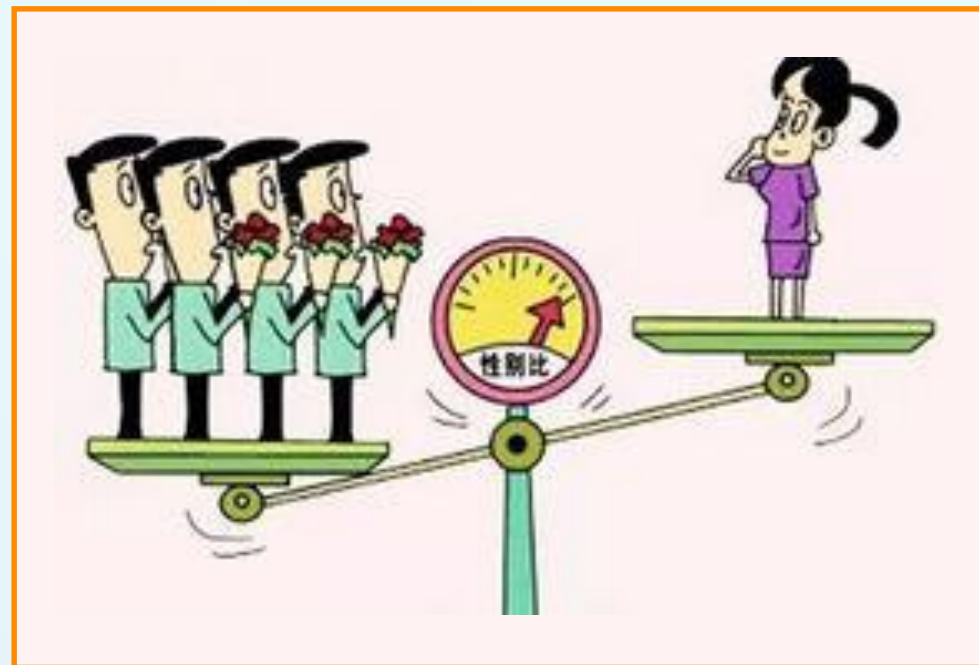
- (1) 提供了一种可广泛应用的，近似计算事件概率的方法。让试验重复大量次数，计算事件 A 发生的频率，用它来近似事件 A 的概率。
- (2) 提供了一种检验理论正确与否的准则。

年份	新生儿总数	男婴数	女婴数	男婴频率	女婴频率
1977	3670	1883	1787	0.5131	0.4869
1978	4250	2177	2073	0.5122	0.4878
1979	4055	2138	1917	0.5273	0.4727
1980	5844	2955	2889	0.5056	0.4944
1981	6344	3271	3073	0.5156	0.4844
1982	7231	3722	3509	0.5147	0.4853
总计	31394	16146	15248	0.5148	0.4852

可以认为生男孩的概率近似值为0.515。 这种概率只能通过统计得出。

中国目前男女出生比例 **117:100**

预计将会出现 **3000万** “光棍”





作业： 6, 7, 9, 10, 11, 13, 14

第

2

讲

谢谢观看