概率论与数理统计

成绩组成:

- 平时成绩60分:

  1.作业和到课情况等25分

  2. 讨论5分(学在浙大 5个讨论题)

  3. 测验30分(学在浙大 3次测验)

- 期末考试40分(设置最低分)

测验范围及测验时间初步安排如下：（可能会有调整）

测验一，内容到第2章
时间：秋学期第7周（周五）21:30-22:30.

测验二：内容到第4章
时间：冬学期第3周（周六）21:30-22:30.

测验三：内容到第7章，
时间：冬学期第6周（周五）21:30-22:30.

- 期末统一考试
- **设置卷面“最低成绩”**
- 最终及格成绩需满足：

总评不低于60分且卷面成绩不低于于“最低成绩”

第一章 概率论的基本概念

- 样本空间，随机事件
- 频率和概率
- 等可能概型
- 条件概率
- 事件的独立性

§1 样本空间，随机事件

自然界与社会生活中的两类现象

$$\begin{cases}\text{确定性现象}\\ \\ \text{不确定性现象}\end{cases}$$

- 确定性现象：结果确定
- 不确定性现象：结果不确定

- 例：
  - 向上抛出的物体会掉落到地上（确定）
  - 打靶，击中靶心（不确定）
  - 买了彩票会中奖（不确定）

概率论与数理统计是研究随机现象数量规律的学科。

这世界唯一确定的就是不确定性，
唯一不变的就是变化

天有不测风云，人有旦夕祸福

江南
鱼戏莲叶东
鱼戏莲叶西
鱼戏莲叶南
鱼戏莲叶北

华罗庚（1910—1985）出生于江苏常州金坛区。数学家，中国科学院院士．他是中国解析数论、矩阵几何学、典型群、自守函数论与多元复变函数论等多方面研究的创始人和开拓者，并被列为芝加哥科学技术博物馆中当今世界88位数学伟人之一．

科学是老老实实的学问，搞科学研究工作就要采取老老实实、实事求是的态度，不能有半点虚假浮夸．

老老实实的态度，首先就是要扎扎实实地打好基础。

要真正打好基础，有两个必经的过程，即“由薄到厚”和“由厚到薄”的过程.“由薄到厚”是学习、接受的过程，“由厚到薄”是消化、提炼的过程.

- 有信心,努力,坚持
- 掌握好基本概念(定义,性质等)
- 多思考,深入思考
- 勤动笔,计算仔细
- 直观想法与严格推导结合
- 一题多解,从不同角度思考
- 举一反三. 条件能放宽?结论能加强?

例：抛硬币出现的正面的频率

试验序号	n =5		n =50		n =500	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	256	0.512
4	5	1.0	25	0.50	253	0.506
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

不确定现象：

$$\begin{cases}\text{个别现象}\\\text{随机现象}\end{cases}$$

在个别试验中其结果呈现出不确定性，但在大量重复试验中其结果又具有统计规律性。

对随机现象的观察、记录、实验统称为随机试验。它具有以下特性：

- 可以在相同条件下重复进行；
- 事先知道可能出现的结果；
- 进行试验前并不知道哪个试验结果会发生。

例：

- 抛一枚硬币，观察试验结果；
- 对89路公交车浙大紫金港校区站登记上车人数；
- 对某批电子产品测试其输入电压；
- 对听课人数进行一次登记；

(一)样本空间

定义：随机试验E的所有结果构成的集合称为E的 **样本空间**，记为S，

称S中的元素e为**样本点**.

- 例：
  - 抛一枚硬币一次,记录结果
  - 记录一城市一日中发生交通事故次数
  - 记录一批产品的寿命x
  - 记录某地一昼夜最高温度x，最低温度y

$S=\{$正面，反面$\}$；

$S=\{0, 1, 2, \cdots\}$；

$S=\{\ x \mid a \leqslant x \leqslant b\ \}$

$S=\{(x, y) \mid T_0 \leqslant y \leqslant x \leqslant T_1\}$；

（二）随机事件

称S的子集A为E的随机事件A，简称事件A. 当且仅当A所包含的一个样本点发生称事件A发生。

- 随机事件有如下特征：

- 事件A是相应的样本空间S的一个子集，其关系可用维恩(Venn)图来表示；
- 事件A发生当且仅当A中的某一个样本点出现；
- 事件A的表示可用集合，也可用语言来表示。

- 例：观察89路公交车浙大紫金港校区站候车人数。

S={0, 1, 2, …}；

A={至少有10人候车}={10, 11, 12, …}

S，A为随机事件，

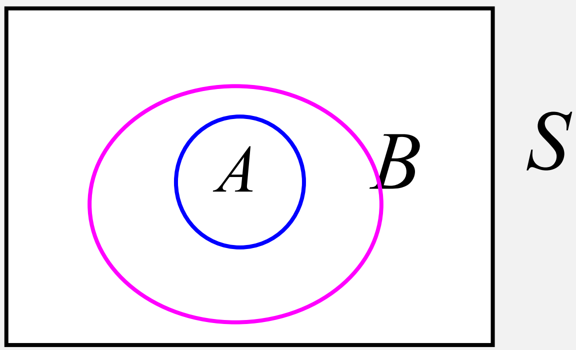
A可能发生，也可能不发生。

- 由一个样本点组成的单点集，称为基本事件。
- 每次试验S总是发生，故又称S为必然事件。
- 记Φ为空集，不包含任何样本点, 则每次试验Φ都不发生, 称Φ为不可能事件。

(三) 事件的关系及运算

❖事件的关系（包含、相等）

1° $A \subset B$: 事件A发生一定导致B发生



$$2^\circ \ A = B \Leftrightarrow \begin{cases} A \subset B \\ B \subset A \end{cases}$$

例：

- 记A={明天天晴}，B={明天无雨} $\Rightarrow B \supset A$
- 记A={至少有10人候车}，B={至少有5人候车}

$$\Rightarrow B \supset A$$

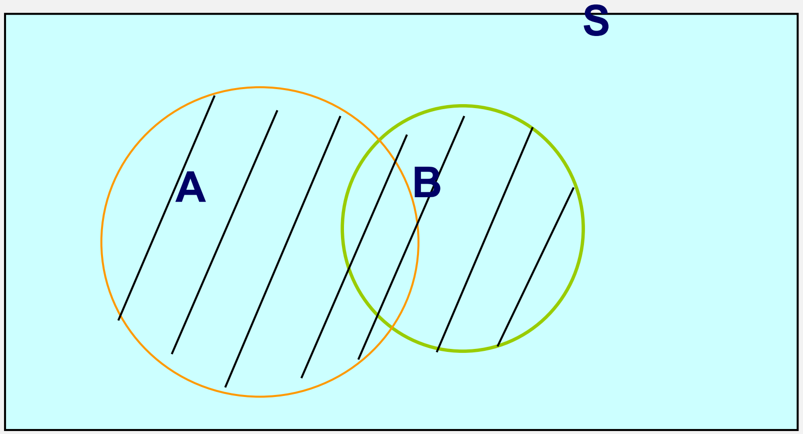
- 抛两颗均匀的骰子，两颗骰子出现的点数分别记为x, y. 记A={x+y为奇数}，B={两次的骰子点数奇偶性不同}，则 $\Rightarrow B = A$

事件的运算

- A与B的和事件，记为 $A\cup B$

$$A\cup B=\{\,x\mid x\in A\ 或\ x\in B\,\}:$$

A与B至少有一发生。

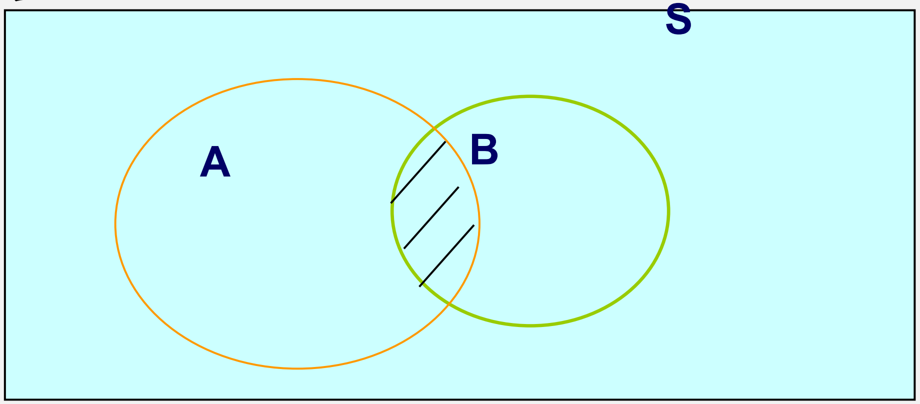


事件的运算

- A与B的积事件，记为$A \cap B, A \cdot B, AB$

$$A \cap B = \{\, x \mid x \in A \text{ 且 } x \in B \,\}:$$

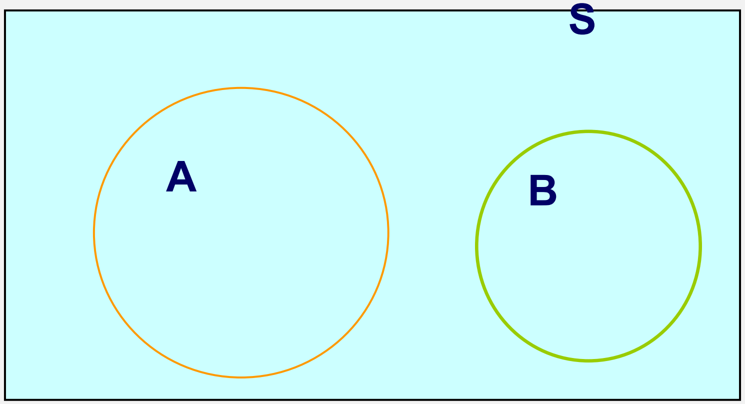
A与B同时发生。



$\bigcup_{i=1}^{n} A_i$：$A_1, A_2, \cdots A_n$至少有一发生

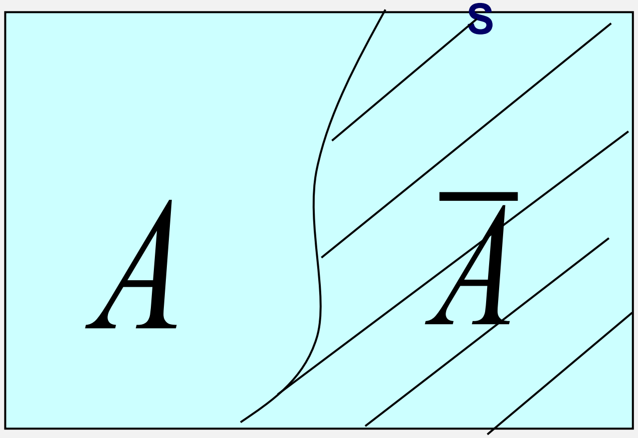
$\bigcap_{i=1}^{n} A_i$：$A_1, A_2, \cdots A_n$同时发生

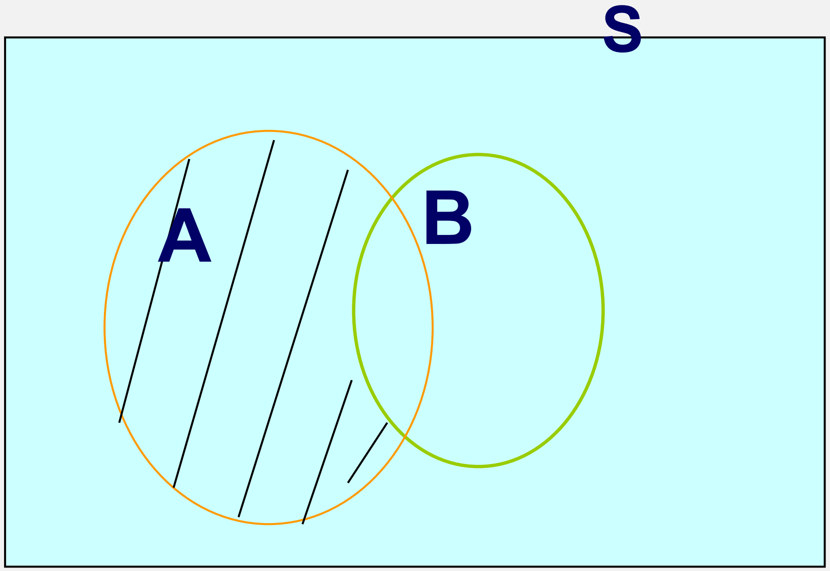
- 当AB= Φ时，称事件A与B是互不相容的，或互斥的.



A的逆事件记为$\overline{A}$, $\begin{cases} A\cup\overline{A}=S \\ A\overline{A}=\varnothing \end{cases}$， 若$\begin{cases} A\cup B=S \\ AB=\varnothing \end{cases}$，

称A, B互逆(互为对立事件)





事件A对事件B的差事件：

$$A\,\overline{B} = A - B = \{\, x \mid x \in A \text{ 且 } x \notin B \,\}$$

“和”、“交”关系式——德摩根定律

$$\overline{\bigcap_{i=1}^{n} A_i} = \bigcup_{i=1}^{n} \overline{A_i} = \overline{A}_1 \cup \overline{A}_2 \cup \cdots \cup \overline{A}_n;$$

$$\overline{\bigcup_{i=1}^{n} A_i} = \bigcap_{i=1}^{n} \overline{A_i} = \overline{A}_1 \overline{A}_2 \cdots \overline{A}_n;$$

例：设A={ 甲来听课 }，B={ 乙来听课 }，则：

$A \bigcup B =$ {甲、乙至少有一人来}

$A \bigcap B =$ {甲、乙都来}

$\overline{A \bigcup B} = \overline{A}\,\overline{B} =$ {甲、乙都不来}

$\overline{A} \bigcup \overline{B} = \overline{AB} =$ {甲、乙至少有一人不来}

概率中常有以下定义：由n个元件组成的系统，其中一个损坏，则系统就损坏，此时这一系统称为“串联系统”；若有一个不损坏，则系统不损坏，此时这一系统称为“并联系统”。

- 例：　由n个部件组成的系统，记

  - 串联系统：$A=\bigcap_{i=1}^{n} A_i$

  - 并联系统：$A=\bigcup_{i=1}^{n} A_i$

A_i ＝{第i个部件没有损坏}，i=1, 2, ⋯ , n,

A＝{系统没有损坏}

§2 频率与概率

(一)频率

定义：记 $f_n(A)=\dfrac{n_A}{n}$;

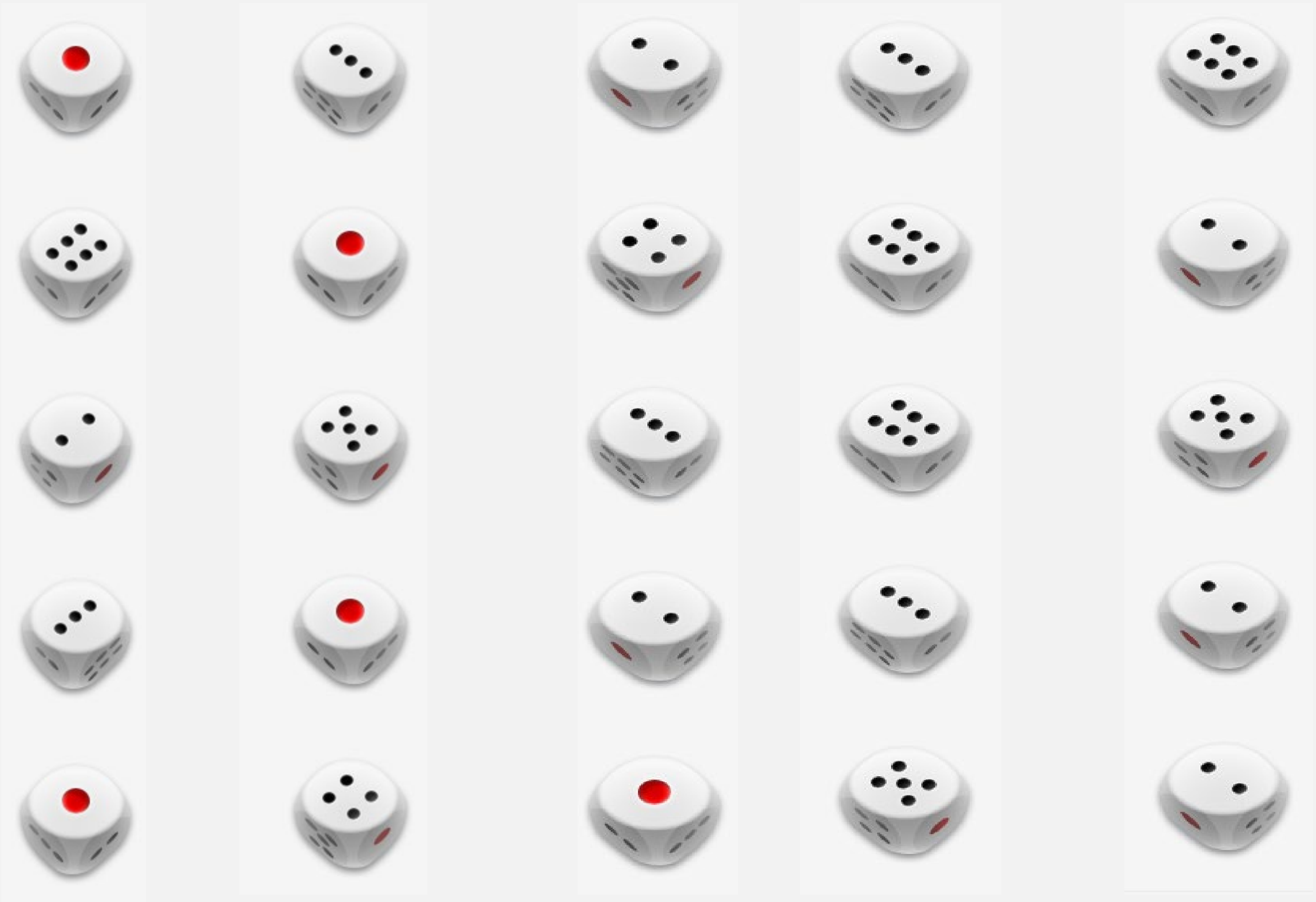
其中 n_A —A发生的次数(频数)；

n—总试验次数。称 $f_n(A)$ 为A在这n次试验中发生的**频率**。

- 某人一共听了16次“概率统计”课，其中有12次迟到，记A={听课迟到}，则

$$f_n(A) = 12/16 = 75\%$$

频率 $f_n(A)$ 反映了事件A发生的频繁程度。



点数	1	2	3	4	5	6
频数	5	6	5	2	3	4
频率	0.2	0.24	0.2	0.08	0.12	0.16

例：2000年悉尼奥运会开幕前，气象学家对两个开幕候选日“9月10日”和“9月15日”的100年气象学资料分析发现，“9月10日”的下雨天数为86天，“9月15日”的下雨天数为22天. 即“9月10日”和“9月15日”的下雨频率分别为86%和22%，因此最后决定开幕日定为“9月15日”。

频率的性质：

1° $0 \le f_n(A) \le 1$

2° $f_n(S) = 1$

3° 若$A_1, A_2, \cdots, A_k$两两互不相容，

则 $f_n\left(\bigcup_{i=1}^{k} A_i\right) = \sum_{i=1}^{k} f_n(A_i)$

例：抛硬币出现的正面的频率

试验序号	n =5		n =50		n =500	
	n_H	$f_n(H)$	n_H	$f_n(H)$	n_H	$f_n(H)$
1	2	0.4	22	0.44	251	0.502
2	3	0.6	25	0.50	249	0.498
3	1	0.2	21	0.42	256	0.512
4	5	1.0	25	0.50	253	0.506
5	1	0.2	24	0.48	251	0.502
6	2	0.4	21	0.42	246	0.492
7	4	0.8	18	0.36	244	0.488
8	2	0.4	24	0.48	258	0.516
9	3	0.6	27	0.54	262	0.524
10	3	0.6	31	0.62	247	0.494

实验者	n	n_H	$f_n(H)$
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

频率的重要性质：

$f_n(A)$ 随$\boldsymbol{n}$的增大渐趋稳定，记稳定值为$\boldsymbol{p}$.

(二) 概率

定义：对样本空间S中任一事件A，定义一个实数$P(A)$，如果满足以下三条：

（1）非负性： $P(A) \geq 0$;

（2）规范性： $P(S) = 1$;

（3）可列可加性：若 $A_1, A_2, \ldots, A_k, \ldots,$ 两两不相容，则 $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i).$

则称$P(A)$为事件A的概率。

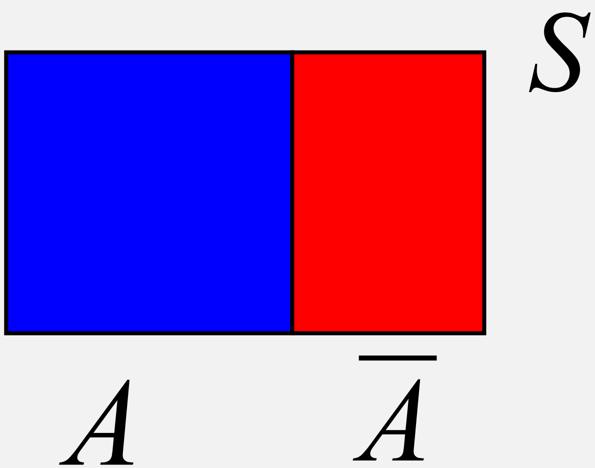
性质：$1^\circ \ P(\varnothing) = 0$

$2^\circ \quad A_1, A_2, \ldots, A_n, A_iA_j = \varnothing, i \neq j,$

$$\Rightarrow P(\bigcup_{i=1}^{n} A_i) = \sum_{i=1}^{n} P(A_i)$$

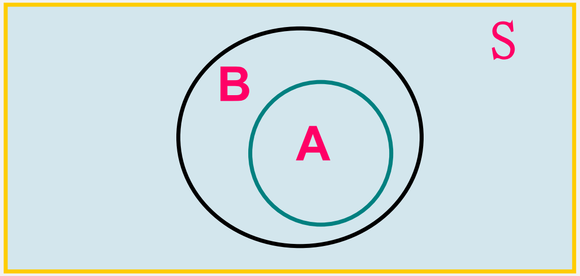
$3^\circ\ P(\overline{A}) = 1 - P(A)$

证：$\because A \cup \overline{A} = S \Rightarrow P(A) + P(\overline{A}) = 1$



4° 若$A \subset B$，则有 $P(B-A)=P(B)-P(A)$

$\Rightarrow P(B) \geq P(A)$, 于是有 $P(A) \leq P(S)=1$

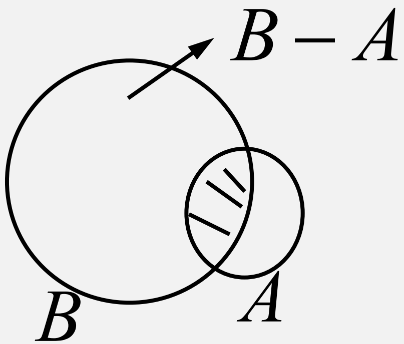


证: $B = A \cup (B-A)$ 不交并

$\Rightarrow P(B)=P(A)+P(B-A)$

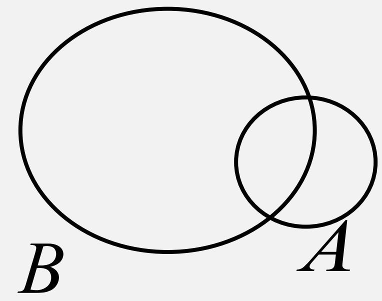
$\Rightarrow P(B-A)=P(B)-P(A)$

问题：一般情况下 $P(B-A)=?$



答案：$P(B-A)=P(B)-P(AB)$

5° 概率的加法公式:



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

证: $\because A \cup B = A \cup (B - A)$

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

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

#5°的推广1：

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

证：$P(A\cup B\cup C)$

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

$$=P(A)+P(B)-P(AB)+P(C)$$

$$-P(AC)-P(BC)+P(ABC)$$

#5°的推广2（一般情形）：

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{1\le i<j\le n} P(A_iA_j)$$

$$+ \sum_{1\le i<j<k\le n} P(A_iA_jA_k) + \cdots + (-1)^{n-1} P(A_1A_2\cdots A_n)$$

例2.1：甲乙丙3人去参加某个集会的概率均为0.4，其中至少有两人参加的概率为0.3，都参加的概率为0.05，求3人中至少有一人参加的概率。

解：设A, B, C分别表示甲, 乙, 丙参加，由条件知

$$P(A) = P(B) = P(C) = 0.4,$$

$$P(AB \cup AC \cup BC) = 0.3,$$

$$P(ABC) = 0.05.$$

由$0.3 = P(AB \cup AC \cup BC) = P(AB)$

$$+ P(AC) + P(BC) - 2P(ABC),$$

得 $$P(AB) + P(AC) + P(BC)$$

$$= 0.3 + 2P(ABC) = 0.4,$$

因此，

P(甲乙丙至少有一人参加）

$= P(A \cup B \cup C)$

$= P(A) + P(B) + P(C) - P(AB)$

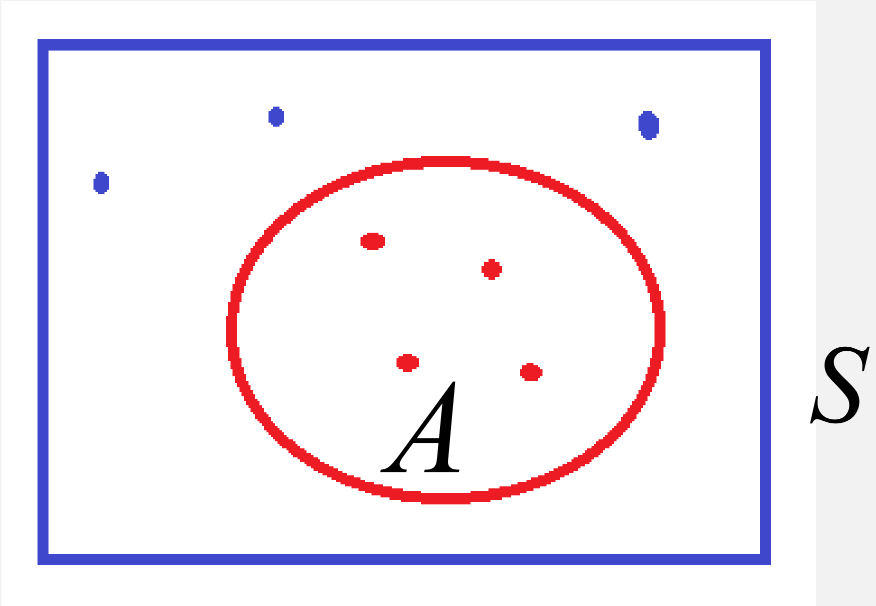
$- P(AC) - P(BC) + P(ABC) = 0.85.$

§3 等可能概型（古典概型）

定义：若试验E满足：

- S中样本点有限(有限性)
- 出现每一样本点的概率相等(等可能性)

称这种试验为等可能概型(或古典概型)。



$$P(A)=\frac{A\text{中样本点个数}}{S\text{中样本点个数}}$$

例3.1：一袋中有8个球，其中3个为红球，5个为黄球，设摸到每一球的可能性相等。

（1）从袋中随机摸一球，记A={ 摸到红球 }，求P(A).

（2）从袋中不放回摸两球，记B={恰是一红一黄}，求P(B).

解：(1)

$$S=\{1,2,\cdots,8\},A=\{1,2,3\}$$

$$\Rightarrow P(A)=\frac{3}{8}$$

$$(2)P(B)=C_3^1C_5^1/C_8^2=\frac{15}{28}\approx 53.6\%$$

例3.2：有N件产品，其中D件是次品，从中不放回的取n件，记A_k＝{恰有k件次品}（k≤D)，求$P(A_k)$.

$$(D \le N, n \le N)$$

解：

$$P(A_k)=C_D^k C_{N-D}^{n-k}/C_N^n,\ k=0,1,\cdots,n$$

（注：当L>m 或 L<0时，记 $C_m^L=0$）

例3.3：将n个不同的球，投入N个不同的盒中(n≤N)，设每一球落入各盒的概率相同，且各盒可放的球数不限，记A＝{ 恰有n个盒子各有一球 }，求P(A).

解： A:"每盒至多一球"

$$P(A)=\frac{N(N-1)(N-2)\ldots(N-n+1)}{N^n}$$

$$=\frac{A_N^n}{N^n}$$

- 应用（生日问题）在一个n（≤365）人的班级里，至少有两人生日相同的概率是多少？

解：

记B＝{至少两人生日相同}

$$则P(B)=1-\frac{A_{365}^{n}}{365^{n}}$$

- 当$n=64$时，$p=0.997$
- 当$n=100$时，$p=0.9999997$

N=365

人数n	(N+1-n)/N	n人生日全不同概率	n人至少两人生日相同概率
1	1	1	0
2	0.997260274	0.997260274	0.002739726
3	0.994520548	0.991795834	0.008204166
4	0.991780822	0.983644088	0.016355912
5	0.989041096	0.972864426	0.027135574
6	0.98630137	0.959537516	0.040462484
7	0.983561644	0.943764297	0.056235703
8	0.980821918	0.925664708	0.074335292
9	0.978082192	0.905376166	0.094623834
10	0.975342466	0.883051822	0.116948178
11	0.97260274	0.858858622	0.141141378
12	0.969863014	0.832975211	0.167024789
13	0.967123288	0.805589725	0.194410275
14	0.964383562	0.776897488	0.223102512
15	0.961643836	0.74709868	0.25290132
16	0.95890411	0.716395995	0.283604005
17	0.956164384	0.684992335	0.315007665
18	0.953424658	0.653088582	0.346911418
19	0.950684932	0.620881474	0.379118526
20	0.947945205	0.588561616	0.411438384
21	0.945205479	0.556311665	0.443688335
22	0.942465753	0.524304692	0.475695308
23	0.939726027	0.492702766	0.507297234
24	0.936986301	0.461655742	0.538344258
25	0.934246575	0.431300296	0.568699704
26	0.931506849	0.40175918	0.59824082
27	0.928767123	0.373140718	0.626859282
28	0.926027397	0.345538528	0.654461472
29	0.923287671	0.319031463	0.680968537
30	0.920547945	0.293683757	0.706316243
31	0.917808219	0.269545366	0.730454634
32	0.915068493	0.246652472	0.753347528
33	0.912328767	0.225028146	0.774971854
34	0.909589041	0.204683135	0.795316865
35	0.906849315	0.185616761	0.814383239
36	0.904109589	0.167817894	0.832182106
37	0.901369863	0.151265992	0.848734008
38	0.898630137	0.135932179	0.864067821
39	0.895890411	0.121780336	0.878219664
40	0.893150685	0.10876819	0.89123181
41	0.890410959	0.096848389	0.903151611
42	0.887671233	0.085969528	0.914030472
43	0.884931507	0.076077144	0.923922856
44	0.882191781	0.067114631	0.932885369
45	0.879452055	0.059024101	0.940975899
46	0.876712329	0.051747157	0.948252843
47	0.873972603	0.045225597	0.954774403
48	0.871232877	0.039402027	0.960597973
49	0.868493151	0.034220391	0.965779609
50	0.865753425	0.02962642	0.97037358
51	0.863013699	0.025568007	0.974431993
52	0.860273973	0.021995491	0.978004509
53	0.857534247	0.018861887	0.981138113
54	0.854794521	0.016123037	0.983876963
55	0.852054795	0.013737711	0.986262289
56	0.849315068	0.011667645	0.988332355
57	0.846575342	0.009877541	0.990122459
58	0.843835616	0.008335021	0.991664979
59	0.84109589	0.007010552	0.992989448
60	0.838356164	0.005877339	0.994122661
61	0.835616438	0.004911201	0.995088799
62	0.832876712	0.004090425	0.995909575
63	0.830136986	0.003395613	0.996604387
64	0.82739726	0.002809521	0.997190479

68	0.816438356	0.001273609	0.998726391
69	0.81369863	0.001036334	0.998963666
70	0.810958904	0.000840424	0.999159576
71	0.808219178	0.000679247	0.999320753
72	0.805479452	0.000547119	0.999452881
73	0.802739726	0.000439194	0.999560806
74	0.8	0.000351356	0.999648644
75	0.797260274	0.000280122	0.999719878
76	0.794520548	0.000222563	0.999777437
77	0.791780822	0.000176221	0.999823779
78	0.789041096	0.000139045	0.999860955
79	0.78630137	0.000109332	0.999890668
80	0.783561644	8.56681E-05	0.999914332
81	0.780821918	6.68915E-05	0.999933109
82	0.778082192	5.20471E-05	0.999947953
83	0.775342466	4.03543E-05	0.999959646
84	0.77260274	3.11779E-05	0.999968822
85	0.769863014	2.40027E-05	0.999975997
86	0.767123288	1.8413E-05	0.999981587
87	0.764383562	1.40746E-05	0.999985925
88	0.761643836	1.07198E-05	0.99998928
89	0.75890411	8.13533E-06	0.999991865
90	0.756164384	6.15164E-06	0.999993848
91	0.753424658	4.6348E-06	0.999995365
92	0.750684932	3.47927E-06	0.999996521
93	0.747945205	2.60231E-06	0.999997398
94	0.745205479	1.93925E-06	0.999998061
95	0.742465753	1.43983E-06	0.99999856
96	0.739726027	1.06508E-06	0.999998935
97	0.736986301	7.84949E-07	0.999999215
98	0.734246575	5.76346E-07	0.999999424
99	0.731506849	4.21601E-07	0.999999578
100	0.728767123	3.07249E-07	0.999999693

例3.4：(抽签问题)一袋中有a个红球，b个白球，记a＋b＝n．设每次摸到各球的概率相等，每次从袋中摸一球，不放回地摸n次。求第k次摸到红球的概率。

记$A_k = \{$第k次摸到红球$\}$, 求$P(A_k)$.

将n个球依次编号为:

$$1,2,\ldots,n$$

其中前a号球是红球

解1:

视1,2,...,n的每一个排列为一样本点,

则每一样本点等概率

$$P(A_k)=\frac{a(a+b-1)!}{(a+b)!}=\frac{a}{a+b}$$

--------与k无关

解2：视哪几次摸到红球为一样本点

每点出现的概率相等

$$\therefore P(A_k) = C_{n-1}^{a-1} / C_n^a = \frac{a}{a+b}$$

解3：将第k次摸到的球号作为一样本点，

则取到各球的概率相等

$$S=\{1,2,\ldots,n\}$$

$$A_k=\{1,2,\ldots,a\}$$

$$\Rightarrow P(A_k)=\frac{a}{n}=\frac{a}{a+b}$$

已算得: 对$1 \le k \le n$都有$P(A_k) = P(A_1) = \frac{a}{n}$.

事实上:

对任何$1 \le m \le a$,任何$1 \le i_1 < i_2 < \cdots < i_m \le n$,有:

$$P(A_{i_1} A_{i_2} \cdots A_{i_m}) = P(A_1 A_2 \cdots A_m)$$

$$= \frac{a(a-1)\cdots(a-m+1)}{n(n-1)\cdots(n-m+1)}.$$

想想为什么呢?

第一次作业:

习题一: 2,3,6,9,10

例3.5：（配对问题）一个小班有n个同学，编号为1, 2, …, n号，中秋节前每人准备一件礼物，相应编号为1, 2, … ,n。将所有礼物集中放在一起，然后每个同学随机取一件，求没有人拿到自己礼物的概率。

解：设A_i表示第i人拿到自己的礼物，i=1,2,…,n， A表示至少有一人拿到自己的礼物。

$$P(A)=P(A_1\cup\ldots\cup A_n)$$

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

$$P(A_i)=\frac{(n-1)!}{n!}=\frac{1}{n},\quad i\text{ 共}n\text{项},$$

$$P(A_iA_j)=\frac{(n-2)!}{n!},\ i<j\text{有}C_n^2\text{项}$$

$$P(A_iA_jA_k)=\frac{(n-3)!}{n!},\ i<j<k\text{有}C_n^3\text{项}$$

$$\vdots$$

$$P(A_1A_2\cdots A_n)=\frac{1}{n!}$$

$$P(\text{没有人取到自己礼物}) = P(\overline{A})$$

$$= 1 - P(A_1 \cup \ldots \cup A_n)$$

$$= 1 - \sum_{i=1}^{n} \frac{1}{n} + C_n^2 \frac{1}{n(n-1)} -$$

$$C_n^3 \frac{1}{n(n-1)(n-2)} + \ldots + (-1)^n \frac{1}{n!}$$

$$= 1 - 1 + \frac{1}{2!} - \frac{1}{3!} + \ldots + (-1)^n \frac{1}{n!}$$

$$= \sum_{i=0}^{n} \frac{(-1)^i}{i!} \approx e^{-1} \approx 0.368$$ 当n很大时

n人	(-1)^n/n!	没人拿到自己礼物概率		1/e
1	-1	0		0.367879441
2	0.5	0.5		
3	-0.166666667	0.333333333		
4	0.041666667	0.375		
5	-0.008333333	0.366666667		
6	0.001388889	0.368055556		
7	-0.000198413	0.367857143		
8	2.48016E-05	0.367881944		
9	-2.75573E-06	0.367879189		
10	2.75573E-07	0.367879464		
11	-2.50521E-08	0.367879439		
12	2.08768E-09	0.367879441		
13	-1.6059E-10	0.367879441		
14	1.14707E-11	0.367879441		
15	-7.64716E-13	0.367879441		
16	4.77948E-14	0.367879441		
17	-2.81146E-15	0.367879441		
18	1.56192E-16	0.367879441		
19	-8.22064E-18	0.367879441		
20	4.11032E-19	0.367879441		

人们在长期的实践中总结得到“概率很小的事件在一次试验中实际上几乎是不发生的”(称之为**实际推断原理**)。

- 例3.6：某接待站在某一周曾接待12次来访，已知所有这12次接待都是在周二和周四进行的，问是否可以推断接待时间是有规定的?

解：假设接待站的接待时间没有规定，而各来访者在一周的任一天中去接待站是等可能的，那么，12次接待来访者都是在周二、周四的概率为

$$2^{12}/7^{12}=0.000\ 000\ 3.$$

现在概率很小的事件在一次试验中竟然发生了，因此，有理由怀疑假设的正确性，从而推断接待站不是每天都接待来访者，即认为其接待时间是有规定的。

§4 条件概率

例4.1：一个家庭中有两个小孩，已知至少一个是女孩，问两个都是女孩的概率是多少？

（假定生男生女是等可能的）

解： 由题意，样本空间为

$S=\{(男，男),(男，女),(女，男),(女，女)\}$

A 表示事件“至少有一个是女孩”，

$A=\{(男，女),(女，男),(女，女)\}$

$B=\{(女，女)\}$

由于事件A已经发生，所以这时试验的所有可能结果只有三种，而事件B包含的基本事件只占其中的一种，所以有

$$P(B|A) = \frac{1}{3}$$

$P(B \mid A)$表示A发生的条件下,B发生的条件概率

在这个例子中，若不知道事件A已经发生的信息，那么事件发生的概率为

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

这里 $P(B)\neq P(B|A)$

其原因在于事件 A 的发生改变了样本空间，使它由原来的 S 缩减为 $S_A=A$ ，而 S_A 是在新的样本空间中由古典概率的计算公式而得到的 $P(B|A)$

例4.2：有一批产品，其合格率为90%，合格品中有95%为优质品，从中任取一件，记A={取到一件合格品}，

B={取到一件优质品}.

则 $P(A)=90\%$ 而$P(B|A)=95\%$.

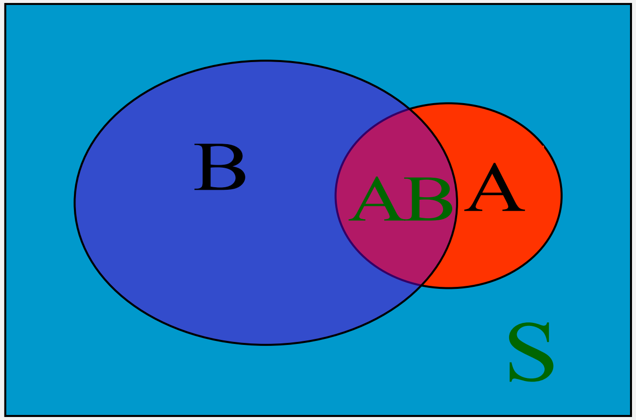
1. P(A)是A在整批产品中所占的概率比例

2. P(B|A)是B在A中所占的概率比例

3. 可将P(A)记为P(A|S)，P(A)也可视为条件概率.

一、条件概率定义

$$P(B|A)=\frac{P(AB)}{P(A)} \qquad P(A)\neq 0$$



性质：$P(.|A)$是概率

(1)非负性：$P(B|A)\geq 0;$

(2)规范性：$P(S|A)=1$

(3)可列可加性：$A_1, A_2, \ldots, A_k, \ldots$，两两互斥

$$\Rightarrow P(\bigcup_{i=1}^{\infty} A_i|A)=\sum_{i=1}^{\infty} P(A_i|A)$$

P(.|A)具有概率的所有性质。如：

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

$$P(B \cup C \mid A) = P(B \mid A) + P(C \mid A) - P(BC \mid A)$$

$$B \supset C \ \Rightarrow P(B \mid A) \geq P(C \mid A)$$

条件概率含义：

设$P(A)>0$.独立重复进行随机试验直到事件A发生为止，记录最后一次结果. 这样的随机试验称为一次新随机试验，对应的概率用P_A表示.

对$B\subset S$,有

$$P_A(B)=\sum_{n=1}^{\infty}P(\text{前}n-1\text{次试验}A\text{未发生，第}n\text{次}A\text{发生且}B\text{发生})$$

$$=\sum_{n=1}^{\infty}[1-P(A)]^{n-1}P(AB)=\frac{P(AB)}{P(A)}=P(B\mid A)$$

也就是新随机试验对应的概率P_A就是条件概率$P(\cdot\mid A)$

例4.3：天气很好，小王想带家人去千岛湖玩，又想到天目山玩．他有一枚硬币，但不知道这枚硬币出现正面的概率．利用这枚硬币设计一个试验帮他做决定，使得最后他去千岛湖和去天目山的概率相等．

解：抛硬币一次，设出现正面的概率是p, 出现反面的概率是$1-p$. 独立抛硬币两次，样本空间

$$S=\{(\text{正},\text{正}),(\text{反},\text{反}),(\text{正},\text{反}),(\text{反},\text{正})\}.$$

令$B_1=\{(\text{正},\text{反})\}, B_2=\{(\text{反},\text{正})\}, B=B_1\cup B_2$.

则$P(B_1)=P(B_2)=p(1-p), P(B_1|B)=P(B_2|B)=1/2$.

于是可以设计这样的试验：

独立重复抛两次硬币，一直到出现结果为(正, 反)或(反, 正)为止. 如果是(正, 反)就去千岛湖，如果是(反, 正)就去天目山.

例：某单位想从8名业务员中等概率地选取一名去外地出差一年. 现有一枚均匀硬币.

利用这枚硬币设计一个试验帮这个单位做决定.

解：对这8名业务员分别编为$1\sim8$号. 注意到抛一次硬币，只能等概率地决定两个结果，所以可以考虑不断地用二分法.

- 先抛一次硬币，如果正面就在前4号里取，否则就在后4号里取，这样范围就缩短到4个号码中.
- 再抛一次硬币，如果正面就在这4号的前2号里取，否则就在这4号的后2号里取. 范围就缩短到2个号码中.
- 最后再抛一次硬币，如果正面就取这2号的前1号，否则就取这2号的后1号.

归纳起来，就是抛硬币三次，样本空间为

$S=\{$正正正, 正正反, 正反正, 正反反,
反正正, 反正反, 反反正, 反反反$\}$.

这是等可能概型.

对应这8个样本点，我们分别取1,2,...,8号.

→ 取3号

正反正

n	2^n
10	1024
11	2048
12	4096
13	8192
14	16384
15	32768
16	65536
17	131072
18	262144
19	524288
20	1048576
21	2097152

n	2^n
22	419 4304
23	838 8608
24	1677 7216
25	3355 4432
26	6710 8864
27	1 3421 7728
28	2 6843 5456
29	5 3687 0912
30	10 7374 1824
31	21 4748 3648
32	42 9496 7296
33	85 8993 4592

全球人口 80亿 左右

例4.4： 某单位想从6名业务员中等概率地选取一名去外地出差一年.现有一枚均匀硬币.利用这枚硬币设计一个试验帮这个单位做决定.

解：对这6名业务员分别编为$1\sim 6$号.

如能实现$1,2,\ldots,8$中等概率取一个数, 则条件概率$P(\cdot\,|\,\{1,2,\ldots,6\})$为等概率的在$1,2,\ldots,6$中取一个数.

于是，可以这样试验：

独立重复抛三次硬币, 直到结果不是“反反正”和“反反反”, 因此最后结果是：正正正，正正反, 正反正, 正反反, 反正正, 反正反, 中的某一个, 对应的取$1,2,\ldots,6$号.

二、乘法公式

当下面的条件概率都有意义时：

$$P(AB)=P(A)\cdot P(B\mid A)=P(B)\cdot P(A\mid B)$$

$$P(ABC)=P(A)P(B\mid A)P(C\mid AB)$$

$$P(A_1 A_2 \cdots A_n)$$

$$= P(A_1)P(A_2 \mid A_1)P(A_3 \mid A_1 A_2)$$

$$\cdots P(A_n \mid A_1 \cdots A_{n-1})$$

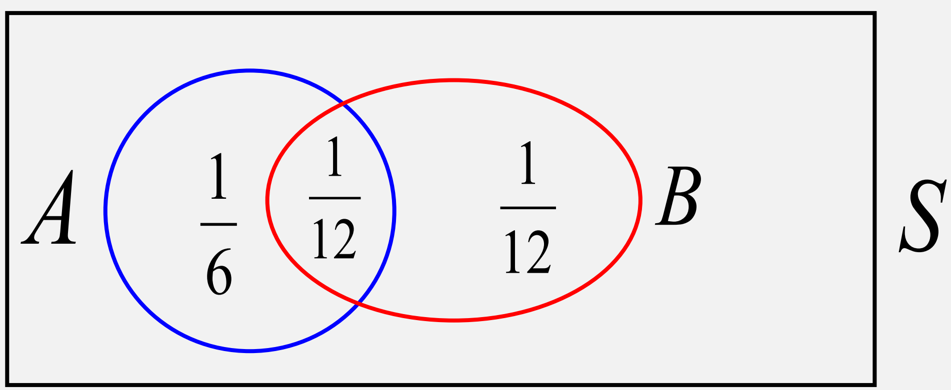
例4. 5: $P(A)=1/4, P(B|A)=1/3,$

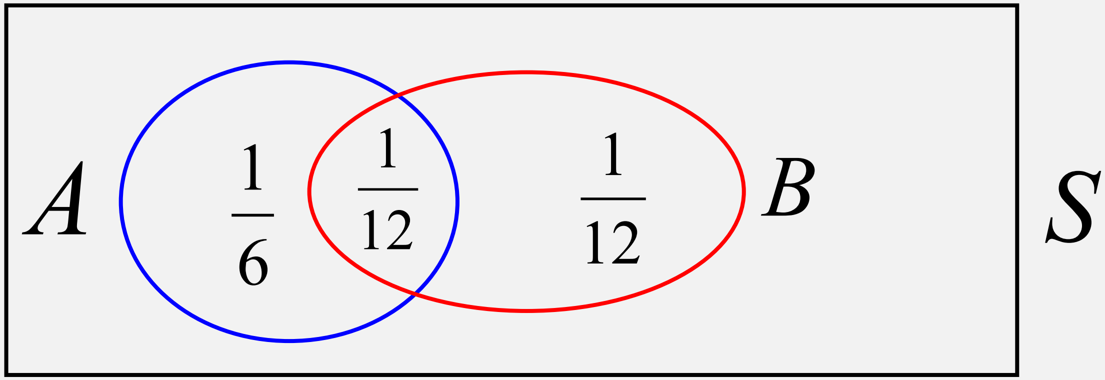
$P(A|B)=1/2,$ 求 $P(A\cup B), P(\bar{A}|A\cup B)$

解： $P(AB)=P(A)P(B|A)=1/12$

$$P(AB)=P(A|B)P(B)\Rightarrow P(B)=1/6$$

于是:





所以：$P(A \cup B) = \frac{1}{6} + \frac{1}{12} + \frac{1}{12} = \frac{1}{3}$

$$P(\overline{A} \mid A \cup B) = 1 - P(A \mid A \cup B) = 1 - \frac{1/4}{1/3} = \frac{1}{4}$$

例4.6：一盒中有5个红球，4个白球，采用不放回抽样，每次取一个，取4次，(1)已知前两次中至少有一次取到红球，求前两次中恰有一次取到红球的概率；(2)已知第4次取到红球，求第1，2次也取到红球的概率。

解：A_i表示第i次取到红球，i=1,2,3,4，B表示前两次中有一次取到红球，C表示前两次中恰有一次取到红球的概率。则

$$P(C|B)=\frac{P(BC)}{P(B)}=\frac{P(C)}{1-P(\bar{B})}=\frac{C_4^1C_5^1/C_9^2}{1-C_4^2/C_9^2}=\frac{2}{3}$$

$$P(A_1A_2|A_4)=\frac{P(A_1A_2A_4)}{P(A_4)}=\frac{P(A_1A_2A_3)}{P(A_1)}$$

$$=\frac{C_5^3/C_9^3}{5/9}=\frac{3}{14}$$

- 例4.7：某厂生产的产品能直接出厂的概率为70%，余下的30%的产品要调试后再定，已知调试后有80%的产品可以出厂，20%的产品要报废。求该厂产品的报废率。

解：设 A={生产的产品要报废}

B={生产的产品要调试}

已知P(B)=0.3，P(A|B)=0.2，

$$A \subset B, A = AB,$$

$$P(A) = P(AB)$$

$$= P(B)P(A|B) = 0.3 \times 0.2 = 6\%$$

例4.8：某行业进行专业劳动技能考核，一个月安排一次，每人最多参加3次；某人第一次参加能通过的概率为60%；如果第一次未通过就去参加第二次，这时能通过的概率为80%；如果第二次再未通过，则去参加第三次，此时能通过的概率为90%。求这人能通过考核的概率。

解：设 A_i={ 这人第i次通过考核 }，

i=1, 2, 3

A={ 这人通过考核 }，

$$A = A_1 \cup \overline{A}_1 A_2 \cup \overline{A}_1 \overline{A}_2 A_3$$

$$P(A)=P(A_1)+P(\overline{A}_1 A_2)+P(\overline{A}_1\overline{A}_2 A_3)$$

$$=P(A_1)+P(\overline{A}_1)\cdot P(A_2\mid\overline{A}_1)$$

$$+P(\overline{A}_1)\cdot P(\overline{A}_2\mid\overline{A}_1)P(A_3\mid\overline{A}_1\overline{A}_2)$$

$$=0.60+0.4\times0.8+0.4\times0.2\times0.9$$

$$=0.992$$

$$\boxed{\begin{aligned}&P(\overline{A}_2\mid\overline{A}_1)\\&=1-P(A_2\mid\overline{A}_1)\\&=1-0.8=0.2\end{aligned}}$$

亦可：

$$P(A)=1-P(\overline{A})=1-P(\overline{A_1}\,\overline{A_2}\,\overline{A_3})$$

$$=1-P(\overline{A_1})P(\overline{A_2}\mid\overline{A_1})P(\overline{A_3}\mid\overline{A_1}\,\overline{A_2})$$

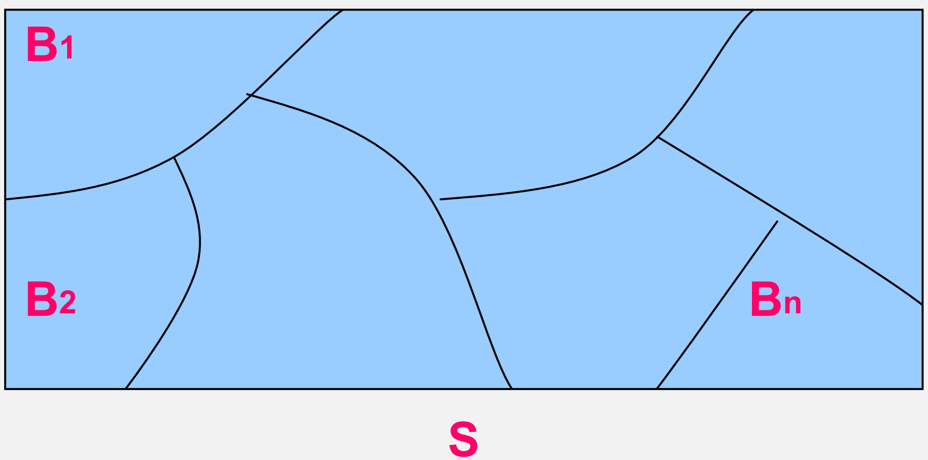
$$=1-0.4\times0.2\times0.1=0.992$$

三、全概率公式与Bayes公式

定义： 称$B_1, B_2, \cdots, B_n$为S的一个**划分**若：

$$(i)\ \text{不漏}\ B_1 \cup B_2 \cup \cdots \cup B_n = S$$

$$(ii)\ \text{不重}\ B_i B_j = \varnothing,\quad i \neq j$$



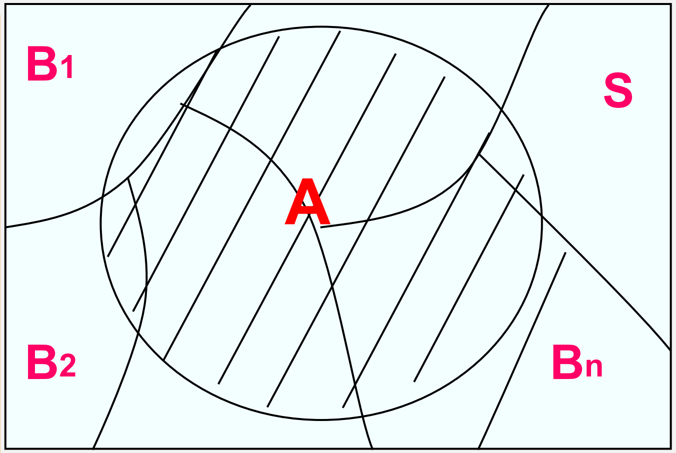
定理：

设$B_1,B_2,\ldots,B_n$为样本空间S的一个划分，

$P(B_i)>0$，$i=1,2,\ldots,n$；则称：

$$P(A)=\sum_{j=1}^{n}P(B_j)\cdot P(A\mid B_j)$$

为全概率公式



证明 $\because A = AS = AB_1 \cup AB_2 \cup \cdots \cup AB_n$

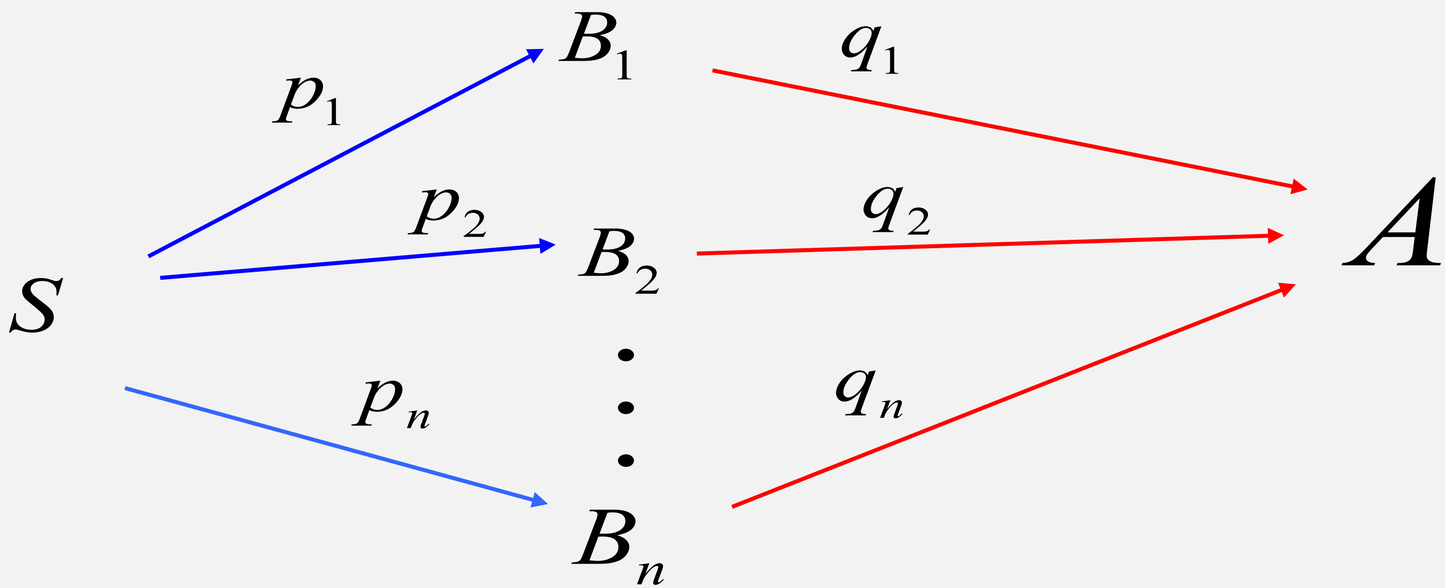
$$\therefore P(A) = \sum_{j=1}^{n} P(AB_j)$$

AB_i与AB_j
不相容$(i \neq j)$

$$= \sum_{j=1}^{n} P(B_j) \cdot P(A \mid B_j)$$

注：在运用全概率公式时，一个关键是构造一组合适的划分。

设 $P(B_j)=p_j, P(A\mid B_j)=q_j, j=1,2,\ldots,n$



$$\boxed{\text{则 } P(A)=\sum_{j=1}^{n} p_j q_j}$$

定理：接上面全概率公式的条件，

且P(A)>0，则

$$P(B_i \mid A)=\frac{P(B_i)P(A \mid B_i)}{\sum_{j=1}^{n} P(B_j)P(A \mid B_j)}$$

称此式为**Bayes公式**。

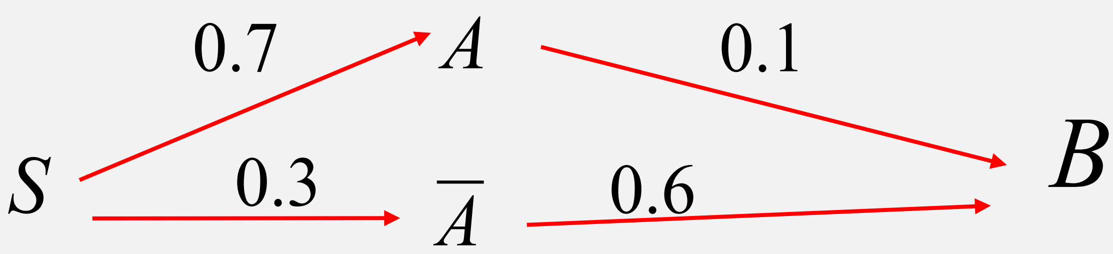
例4.9：一单位有甲、乙两人，已知甲近期出差的概率为70%.若甲出差，则乙出差的概率为10%；若甲不出差，则乙出差的概率为60%。

(1)求近期乙出差的概率；

(2)若已知乙近期出差在外，求甲出差的概率。

解：设A={甲出差}，B={乙出差}

已知 $P(A)=0.70,\quad P(B\mid A)=0.10,\quad P(B\mid \bar{A})=0.60$



(1)由全概率公式：

$$P(B)=P(A)P(B\mid A)+P(\bar{A})P(B\mid \bar{A})$$

$$=0.7\times 0.1+0.3\times 0.6=25\%$$

(2)由Bayes公式：

$$P(A\mid B)=\frac{P(AB)}{P(B)}=\frac{P(A)P(B\mid A)}{P(B)}=\frac{7}{25}$$

例4.10：根据以往的临床记录，某种诊断癌症的试验具有5%的假阳性及5%的假阴性：即

设A={试验反应是阳性}，C={被诊断患有癌症}

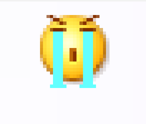
则有：$P(A \mid \overline{C}) = 5\%, P(\overline{A} \mid C) = 5\%.$ 已知某一群体P(C)=0.005，问这种方法能否用于普查？

解：

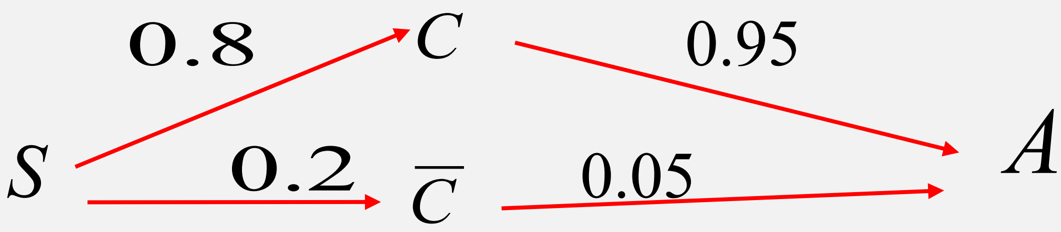
$$\begin{tikzcd} & C \arrow[r, "0.95"] & A \\ S \arrow[ur, "0.005"] \arrow[r, "0.995"] & \overline{C} \arrow[ur, "0.05"] & \end{tikzcd}$$

由 Bayes公式：

$$P(C\mid A)=\frac{P(C)\cdot P(A\mid C)}{P(C)P(A\mid C)+P(\overline{C})P(A\mid \overline{C})}=0.087$$

若用于普查，100个阳性病人中被诊断患有癌症的大约有8.7个，所以不宜用于普查。

若$P(C)$很大，比如$P(C)=0.8$,则



$$P(C\mid A)=\frac{0.8\times 0.95}{0.8\times 0.95+0.2\times 0.05}=0.987$$

说明此方法在医院可用

§5　独立性

例：有10件产品，其中8件为正品，2件次品。从中取2次,每次取1件，设A_i={第i次取到正品}，i=1,2

不放回抽样时，$P(A_2 \mid A_1)=\dfrac{7}{9}\neq P(A_2)=\dfrac{8}{10}$

放回抽样时，$P(A_2 \mid A_1)=\dfrac{8}{10}=P(A_2)$

即放回抽样时，A_1的发生对A_2发生的概率不影响。

定义：设A，B为两随机事件，如果P(AB)=P(A)*P(B)，则称A，B 相互独立.

若$P(A) \neq 0$,则

P(AB)=P(A)P(B)等价于P(B|A)=P(B).

A,B相互独立 $\Leftrightarrow$ $A,\overline{B}$相互独立

$\Leftrightarrow$ $\overline{A},B$相互独立 $\Leftrightarrow$ $\overline{A},\overline{B}$相互独立

证: $P\left(A\overline{B}\right)=P\left(A-AB\right)=P\left(A\right)-P\left(AB\right)$

$\therefore$ A,B相互独立 $\Leftrightarrow P\left(AB\right)=P\left(A\right)P\left(B\right)$

$\Leftrightarrow P\left(A\overline{B}\right)=P\left(A\right)\left[1-P\left(B\right)\right]=P\left(A\right)P\left(\overline{B}\right)$

$\Leftrightarrow$ $A,\overline{B}$相互独立

定义： 设$A_1, A_2, \cdots, A_n$为n个随机事件，

若对$2 \le k \le n, 1 \le i_1 < i_2 < \cdots < i_k \le n$

均有：$P\left(A_{i_1} A_{i_2} \cdots A_{i_k}\right) = \prod_{j=1}^{k} P\left(A_{i_j}\right)$

则称$A_1, A_2, \cdots, A_n$相互独立

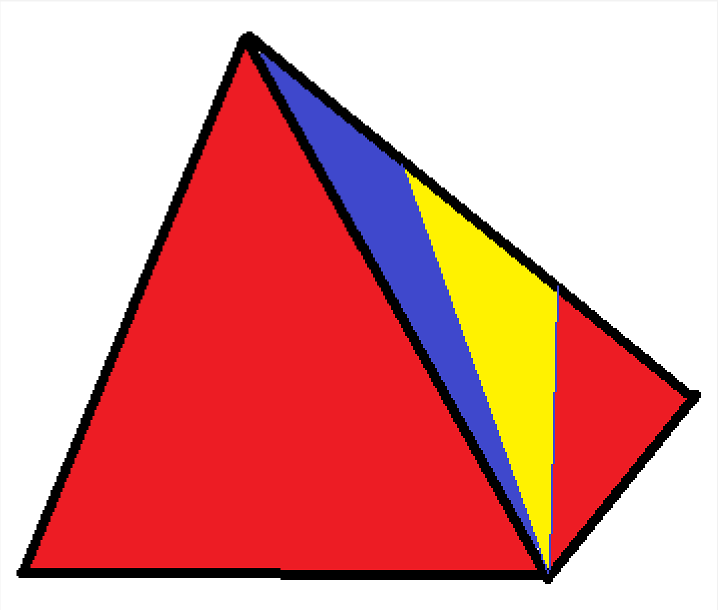
例5.1：有一个正四面体，现在给一面漆上红色，一面漆上黄色，一面漆上蓝色，还有一面漆上红黄蓝三色．现在任取一面.令

$A=$"这面含红色", $B=$"这面含黄色",

$C=$"这面含蓝色"。

问:A,B,C是否两两独立？

是否相互独立？



解：对这四面分别标号为1,2,3,4.

则$S=\{1,2,3,4\}$,

$$A=\{1,4\},B=\{2,4\},C=\{3,4\}$$

$$AB=AC=BC=ABC=\{4\}$$

$$P(A)=P(B)=P(C)=1/2$$

$$P(AB)=P(AC)=P(BC)=P(ABC)=1/4$$

$\therefore \boxed{\text{两两独立}}$,即 $P(AB)=P(A)P(B),$

$$P(AC)=P(A)P(C),$$

$$P(BC)=P(B)P(C)$$

但$\boxed{\text{不是相互独立}}$

$\because P(ABC)\neq P(A)P(B)P(C)$

注意：

1° 相互独立 $\Rightarrow$ 两两独立

2° 两两独立不能 $\Rightarrow$ 相互独立

第二周作业:

习题一：12，13，16，19,25

设一个试验是由一系列子试验组成，

独立试验：指任一次子试验出现的结果都不影响其他各子试验出现的结果；

例如观察十期彩票的开奖结果，是独立试验.

重复试验：如果各子试验是在相同条件下进行的。

例5. 2: $P(A)=0.5, P(B)=0.4$, 求下列情况下$P(A\cup B)$

（1）A与B独立, （2）A与B不相容,

(3) $A\supset B$, (4) $P(AB)=0.3$.

解:(1)$P(A\cup B)=P(A)+P(B)-P(A)P(B)=0.7$

或 $P(A\cup B)=1-P(\overline{A}\overline{B})=1-P(\overline{A})P(\overline{B})=0.7$

(2) $P(A\cup B)=P(A)+P(B)=0.9$

(3) $P(A\cup B)=P(A)=0.5$

(4) $P(A\cup B)=P(A)+P(B)-P(AB)=0.6$

例5.3：甲、乙两人进行乒乓球比赛，每局甲胜的概率为p，$p \geqslant \frac{1}{2}$，问对甲而言，采用三局二胜制有利，还是采用五局三胜制有利？（设各局胜负相互独立）

解：设$A_i=\{$第i局甲胜$\}$

$$\Rightarrow P(A_i)=p,\ i=1,2,\cdots,5$$

再设$A=\{$甲胜$\}$

(1) 三局二胜制：

$$P(A)=P\left(A_1A_2\cup A_1\overline{A}_2A_3\cup\overline{A}_1A_2A_3\right)$$

$$=p^2+2p^2(1-p)\overset{\text{记为}}{=}p_1$$

(2)五局三胜制：

$$P(A)=P\{A_1A_2A_3\cup(\text{前三次有两次赢})A_4$$

$$\cup(\text{前四次有两次赢})A_5\}$$

$$=p^3+C_3^1(1-p)p^3+C_4^2(1-p)^2p^3\overset{\text{记为}}{=}p_2$$

$$p_2 - p_1 = 3p^2(p-1)^2(2p-1)$$

$$\Rightarrow \begin{cases} p_2 > p_1,\ 当p > \dfrac{1}{2} \\ p_2 = p_1,\ 当p = \dfrac{1}{2} \end{cases}$$

例5.4：有5个独立元件构成的系统(如图1)，设每个元件能正常运行的概率为p，求系统正常运行的概率。

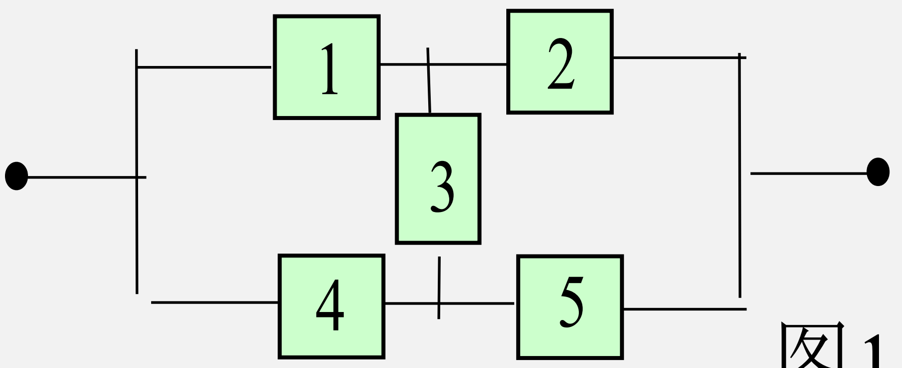
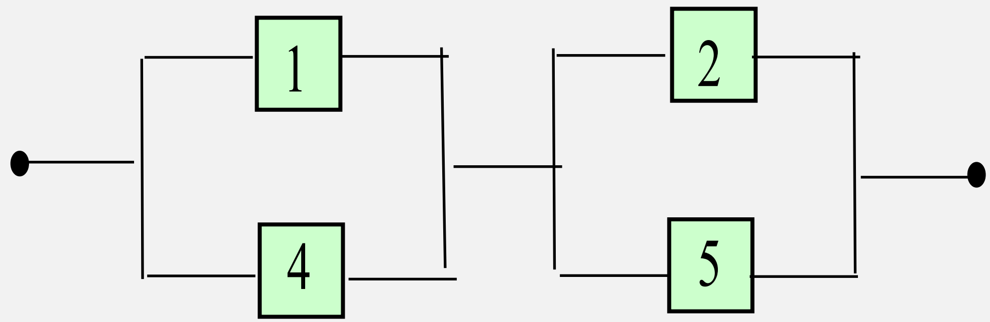


图1

解：设$A_i=\{$第i个元件运行正常$\}, i=1,2,3,4,5$

$$A=\{\text{系统运行正常}\}$$

$$P(A)=P(A_3)\cdot P(A|A_3)+P(\overline{A}_3)\cdot P(A|\overline{A}_3)$$



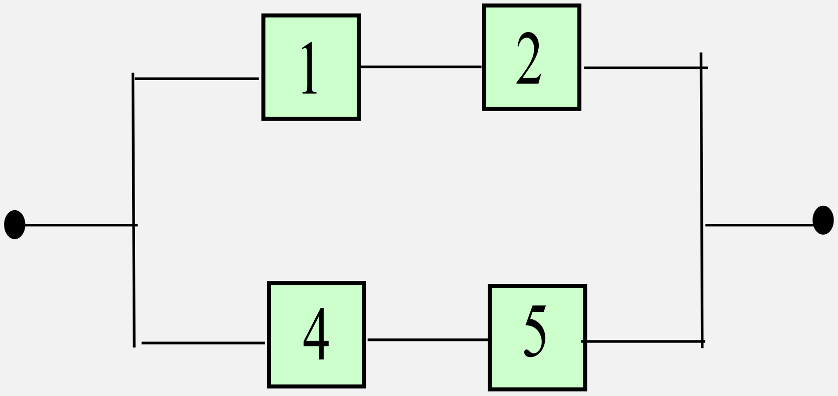
$$p_1 \triangleq P(A|A_3)=P\big((A_1\cup A_4)(A_2\cup A_5)\big)$$

$$=[P(A_1\cup A_4)]^2=(2p-p^2)^2$$

$$p_2 \triangleq P(A \mid \overline{A}_3) = P\left(A_1A_2 \cup A_4A_5\right) = 2p^2 - p^4$$

$$P(A) = p(2p - p^2)^2 + (1-p)(2p^2 - p^4)$$

$$= 2p^2 + 2p^3 - 5p^4 + 2p^5$$



例5.5：一袋中有编号为1,2,3,4共4个球，采用有放回抽样，每次取一球，共取2次，记录号码之和，这样独立重复进行试验，求“和等于3”出现在“和等于5”之前的概率。

解：设A表示“和等于3”出现在“和等于5”之前，

B表示第一次号码之和为3，

C表示第一次号码之和为5，

D表示第一次号码之和既不为3也不为5

$$P(B)=\frac{2}{16},\quad P(C)=\frac{4}{16},\quad P(D)=\frac{10}{16}$$

$$P(A)=P(B)P(A|B)+P(A|C)P(C)+P(A|D)P(D)$$

$$=\frac{2}{16}\times 1+\frac{4}{16}\times 0+\frac{10}{16}\times P(A|D)$$

$$\Rightarrow\quad P(A)=\frac{1}{3}.$$

$$P(A|D)=P(A)$$

在第一次和不等于3或5的情况下求A的条件概率，相当于重新考虑A的概率。

例5.6：某技术工人长期进行某项技术操作，他经验丰富，因嫌按规定操作太过烦琐，就按照自己的方法进行，但这样做有可能发生事故。设他每次操作发生事故的概率为p，$p>0$，但很小很小，他独立重复进行了n次操作，求(1) n次都不发生事故的概率；(2) 至少有一次发生事故的概率。

解：设A={n次都不发生事故},

B={至少有一次发生事故},

C_i={第i次不发生事故},i=1,2,…,n

则$C_1,\ldots,C_n$相互独立，$P(C_i)=1-p$

$$P(A)=P(C_1\ldots C_n)=(1-p)^n$$

$$P(B)=1-P(A)=1-(1-p)^n,$$

注意到$\lim_{n\to\infty} P(B) = 1$

上式的意义为："小概率事件"在大量独立重复试验中"至少有一次发生"几乎是必然的。

例5.7. 设某地每天发生雾霾的概率为0.2. 在雾霾天气, 该地各居民独立地以概率0.2戴口罩, 在没有雾霾的时候各居民独立地以概率0.01戴口罩. 某天

(1) 在该地任选一居民, 求他戴口罩的概率;

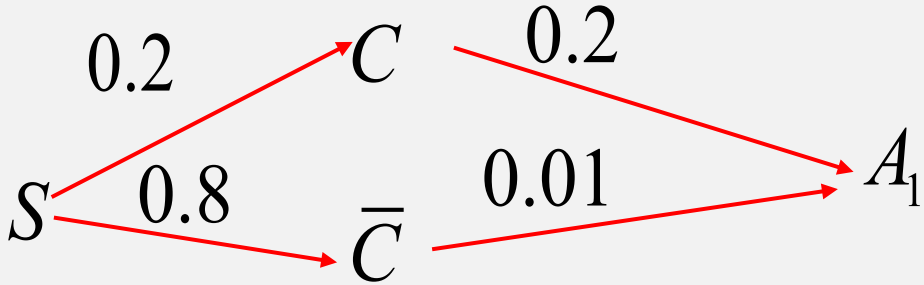
(2) 若选n人, 求他们都戴口罩的概率;

(3) 若选n人发现他们都戴口罩, 求这一天发生雾霾的概率. (这里n为正整数.)

解：

令$C=\{$这一天雾霾$\}, A_i=\{$第i个人带口罩$\}$

(1)



由全概率公式:

$$P(A_1)=P(C)P(A_1|C)+P(\overline{C})P(A_1|\overline{C})=0.048$$

(2) 所求概率为 $P(A_1 A_2 \ldots A_n)$.

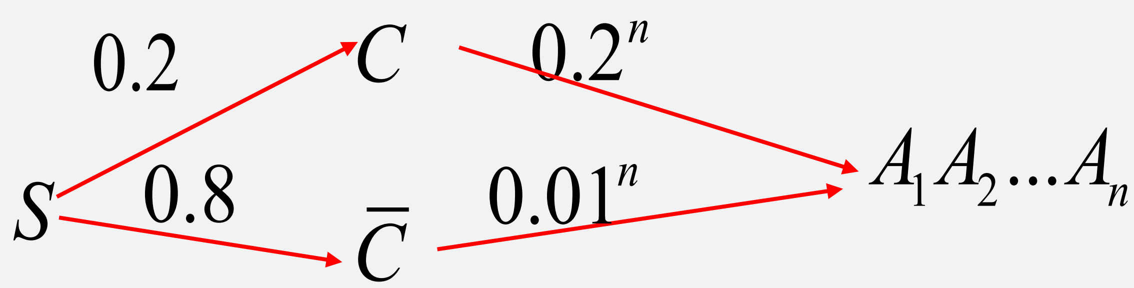
有人认为 $A_1, A_2, \ldots, A_n$ 相互独立, 所以

$P(A_1 A_2 \ldots A_n) = P(A_1) \ldots P(A_n) = 0.048^n$. 对吗?

直观地看, 如果 $A_1, A_2, \ldots, A_{n-1}$ 发生, 也就是前面 $n-1$ 个人都带口罩, 那么那天雾霾的概率就会很大, 从而第 n 个人带口罩的概率也会很大, 也就是 A_n 发生的概率就会变大. 所以 $A_1, A_2, \ldots A_n$ 不太会独立.

举一个特例，如果$P(A_i | C) = 1, P(A_i | \overline{C}) = 0$.也就是只要雾霾天，所有人带口罩；只要不是雾霾天那么没人带口罩．在这种情况下，只要看看一个居民戴口罩的情况，就可以判断是否雾霾天．如果他戴口罩，那么雾霾，从而所有人戴口罩；如果他不戴口罩，那么非雾霾，从而所有人不戴口罩．因此$A_1, A_2, \ldots A_n$不会独立．

那么$P(A_1A_2\dots A_n)$该怎么计算呢？注意到所有人面对的是同样的天气，所以可以按照是否雾霾天作一划分并用全概率公式求得.



由全概率公式：

$$P(A_1A_2\dots A_n)=P(C)P(A_1A_2\dots A_n|C)+P(\bar{C})P(A_1A_2\dots A_n|\bar{C})$$

$$=0.2\times0.2^n+0.8\times0.01^n$$

(3)

$$\begin{tikzcd}
 & C \arrow[dr, "0.2^n"] & \\
S \arrow[ur, "0.2"] \arrow[dr, "0.8"'] & & A_1A_2\ldots A_n \\
 & \bar{C} \arrow[ur, "0.01^n"'] &
\end{tikzcd}$$

由Bayes公式，所求概率为：

$$P(C\mid A_1A_2\ldots A_n)=\frac{P(C)P(A_1A_2\ldots A_n\mid C)}{P(C)P(A_1A_2\ldots A_n\mid C)+P(\bar{C})P(A_1A_2\ldots A_n\mid \bar{C})}$$

$$=\frac{0.2\times 0.2^n}{0.2\times 0.2^n+0.8\times 0.01^n}=\frac{1}{1+4\times 0.05^n},$$ **记为** p_n

则p_n**关于**n**单调递增**，$\lim\limits_{n\to\infty}p_n=1$. **符合直观！**

$$p_1=0.8333,\ p_2=0.9901,\ p_3=0.9995$$

（一）贝叶斯公式介绍

贝叶斯公式，解决的是由果溯因的推理.

假设共有n种两两互斥的原因$B_1, B_2, \ldots, B_n$会导致A发生. 当结果A发生时，我们就会追溯A发生的原因，需要计算由于原因B_j导致A发生的概率是多大？

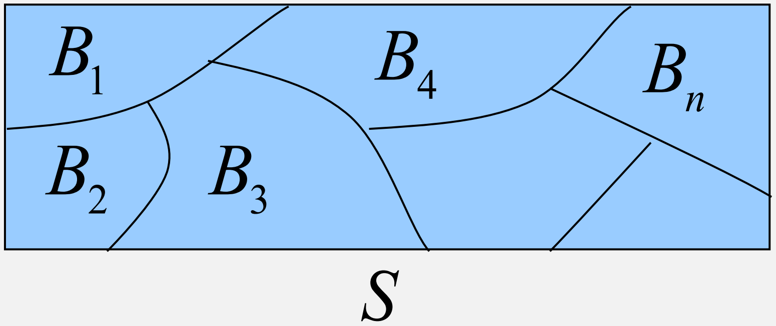
这个概率就是$P(B_j \mid A)$，可由贝叶斯公式给出.

通常，我们会找那个最有可能发生的原因，也就是找B_j，使得$P(B_j \mid A)$是$P(B_1 \mid A), P(B_2 \mid A)..., P(B_n \mid A)$中最大的一个.这个推断方法称之为**贝叶斯方法**.

定义: 称$B_1, B_2, \cdots, B_n$为S的一个划分，若

(i) 不漏 $B_1 \cup B_2 \cup \cdots \cup B_n = S$,

(ii) 不重 $B_i B_j = \varnothing, \ i \neq j.$

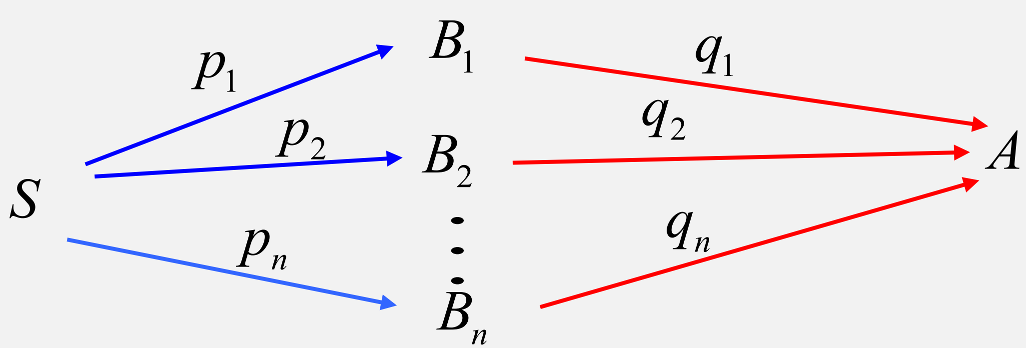


设$B_1,B_2,\cdots,B_n$为S的一个划分且$P(B_i)>0$. 对$P(A)>0$

有**Bayes公式**:

$$P(B_i\mid A)=\frac{P(B_i)P(A\mid B_i)}{\sum_{j=1}^{n}P(B_j)P(A\mid B_j)}=\frac{p_iq_i}{\sum_{j=1}^{n}p_jq_j}$$

设$P(B_j)=p_j, P(A\mid B_j)=q_j, j=1,2,\ldots,n.$



$P(B_i)$先验概率

$P(B_i\mid A)$后验概率

贝叶斯公式由英国数学家托马斯·贝叶斯(1702-1762)提出. 不过贝叶斯在世时并没有公开发表这一重大发现. 而是他去世后两年才由他的朋友理查德·普莱斯整理遗稿时发现并帮助发表的.

贝叶斯方法的应用：

·疾病诊断

·垃圾邮件过滤

·信号检测

·侦破案件

·人工智能

·贝叶斯统计

……

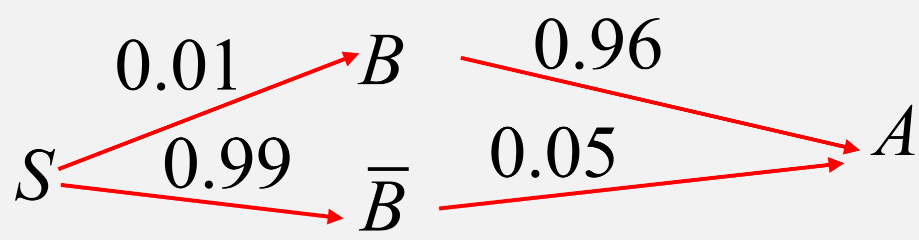
(二)贝叶斯公式的一些应用

例1.(疾病诊断)某种疾病的诊断试验有5%的假阳性和4%的假阴性. 即令$B=\{患有此种疾病\}, A=\{试验反应是阳性\}$,则有$P(A\mid\overline{B})=0.05, P(\overline{A}\mid B)=0.04$.已知此病发病率是0.01.

(1)当试验反应是阳性时,此人患有此种疾病的概率为多少?

(2)为提高准确率,通常会对第一次试验阳性的人再做一次独立的检查.如果这两次都是阳性,问此人患有此种疾病的概率为多少?

解:$(1)B=\{$**患有此种疾病**$\},A=\{$**试验反应是阳性**$\}$

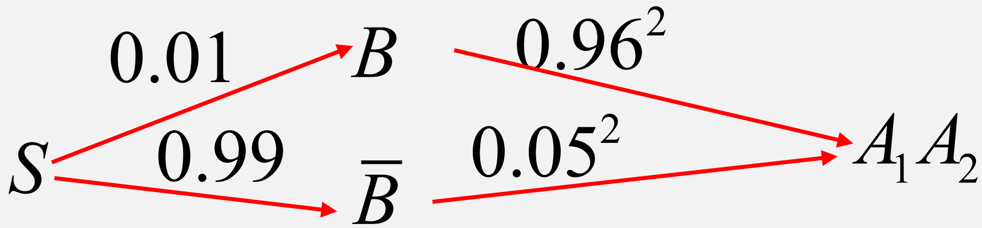


由Bayes公式:

$$P(B\mid A)=\frac{P(B)P(A\mid B)}{P(B)P(A\mid B)+P(\overline{B})P(A\mid\overline{B})}$$

$$=\frac{0.01\times0.96}{0.01\times0.96+0.99\times0.05}=0.1624$$

(2) $B=\{$患有此种疾病$\}$, 令$A_i=\{$第i次试验阳性$\}$,



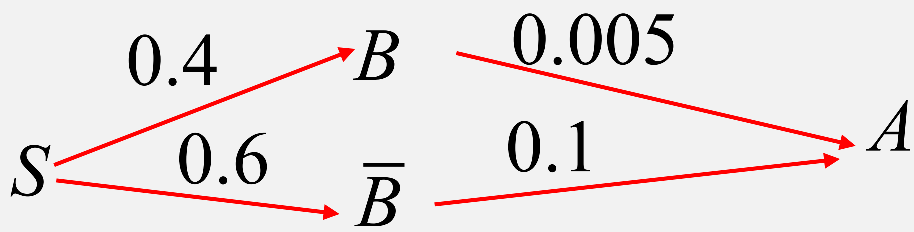
由Bayes公式:

$$P(B\mid A_1A_2)=\frac{P(B)P(A_1A_2\mid B)}{P(B)P(A_1A_2\mid B)+P(\overline{B})P(A_1A_2\mid\overline{B})}$$

$$=\frac{0.01\times0.96^2}{0.01\times0.96^2+0.99\times0.05^2}=0.7883$$

例2.（垃圾邮件过滤）某人的邮箱收到正常邮件的概率为0.4,垃圾邮件的概率为0.6.正常邮件里包含词语“免费”的概率为0.005,垃圾邮件里包含词语“免费”的概率为0.1.现在此人设置把含有词语“免费”的邮件自动过滤到垃圾箱中.问过滤到垃圾箱中的邮件确实是垃圾邮件的概率为多少？

解： 令$A=\{$被过滤到垃圾箱中$\}$，$B=\{$是正常邮件$\}$，



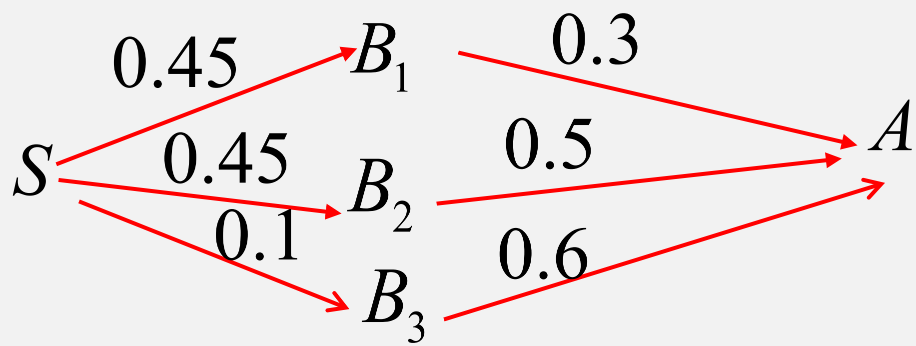
由Bayes公式：

$$P(\overline{B}\mid A)=\frac{P(\overline{B})P(A\mid\overline{B})}{P(B)P(A\mid B)+P(\overline{B})P(A\mid\overline{B})}$$

$$=\frac{0.6\times0.1}{0.4\times0.005+0.6\times0.1}=0.9677$$

例3.（最大后验概率准则）小王参加一个棋类比赛. 其中45%为一类棋手, 小王赢他们的概率为0.3; 45%为二类棋手, 小王赢他们的概率为0.5; 其余为三类棋手, 小王赢他们的概率为0.6. 从这些棋手中任选一人与小王比赛. 如果小王获胜了，你觉得此人最有可能是哪类棋手?

解： 令$B_i = \{$此人是i类棋手$\}$，$i = 1,2,3, A = \{$小王赢$\}$，



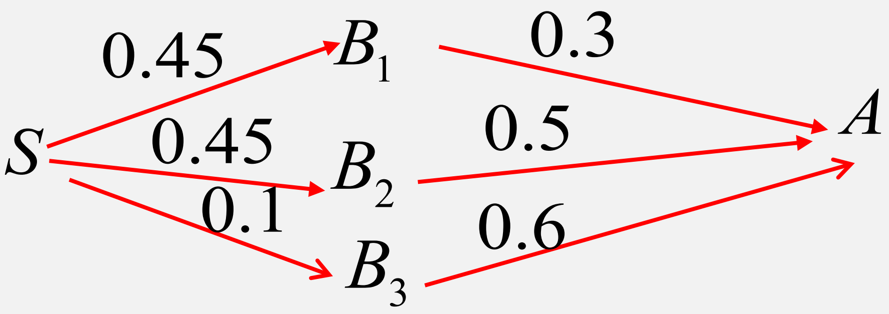
由Bayes公式：

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{P(B_1)P(A \mid B_1) + P(B_2)P(A \mid B_2) + P(B_3)P(A \mid B_3)}$$

$$P(B_1 \mid A) = \frac{0.45 \times 0.3}{0.45 \times 0.3 + 0.45 \times 0.5 + 0.1 \times 0.6} = 0.3214$$

$$P(B_2 \mid A) = \frac{0.45 \times 0.5}{0.45 \times 0.3 + 0.45 \times 0.5 + 0.1 \times 0.6} = 0.5357$$

$$P(B_3 \mid A) = \frac{0.1 \times 0.6}{0.45 \times 0.3 + 0.45 \times 0.5 + 0.1 \times 0.6} = 0.1429$$



$\because P(B_2 \mid A) > P(B_1 \mid A)$

且$P(B_2 \mid A) > P(B_3 \mid A)$,

$\therefore$ 此人最有可能是二类棋手

课件待续!

2023/9/16

THE END