

人工智能顶级实战工程师**就业课程**

数学基础-概率论和数理统计



主讲人：米老师

概率论的基本概念

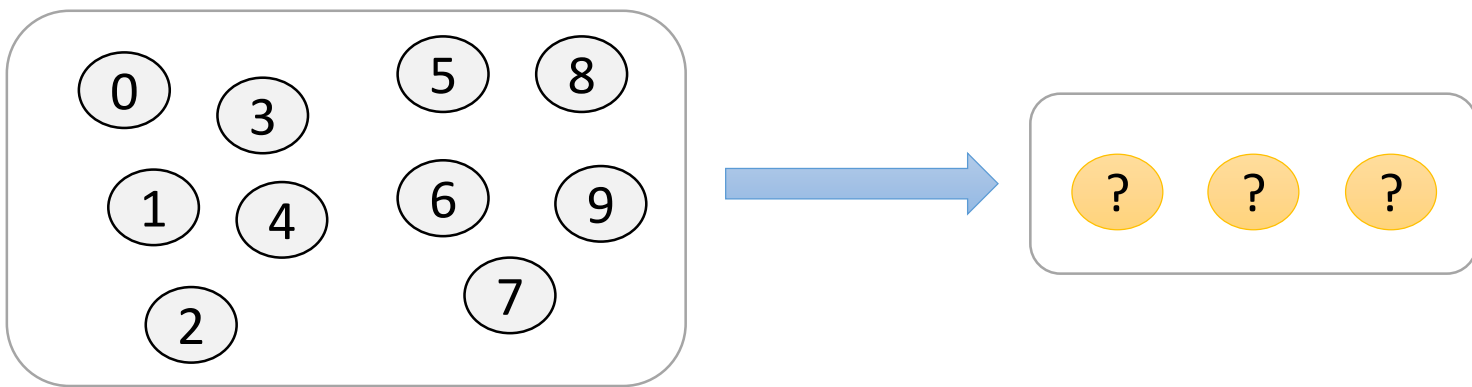
排列数

从 m 个不同元素中取出 n ($n \leq m$) 个元素(被取出的元素各不相同), 并按照一定的顺序排成一行(一般顺序是抽取出来的顺序), 叫做从 m 个不同元素中取出 n 个元素的一个排列。记作: $A(m,n)$

$$A(m, n) = A_m^n = \frac{m!}{(m-n)!}$$

案例

在一个盒子中有十个完全相同的球，其中每个球上编有一个编号，球的编号从0到9，求随机抽取3个球，可能出现的数字序列共有多少种？
（备注：考虑数字的顺序，认为1、2、3和3、2、1是不一样的）。



解：

$$A_{10}^3 = \frac{10!}{(10-3)!} = 720$$

总共10个球，抽取3个球的排列数：

第一步：从10个球中，获取一个球，有10种选择方式

第二步：从剩下的9个球中，获取一个球，有9种选择方式

第三步：从剩下的8个球中，获取一个球，有8种选择方式

合并这3步，就共有 $10 * 9 * 8$ 种选择方式，即

$$A_{10}^3 = 10 * 9 * 8 = \frac{10!}{(10-3)!} = 720$$

组合数

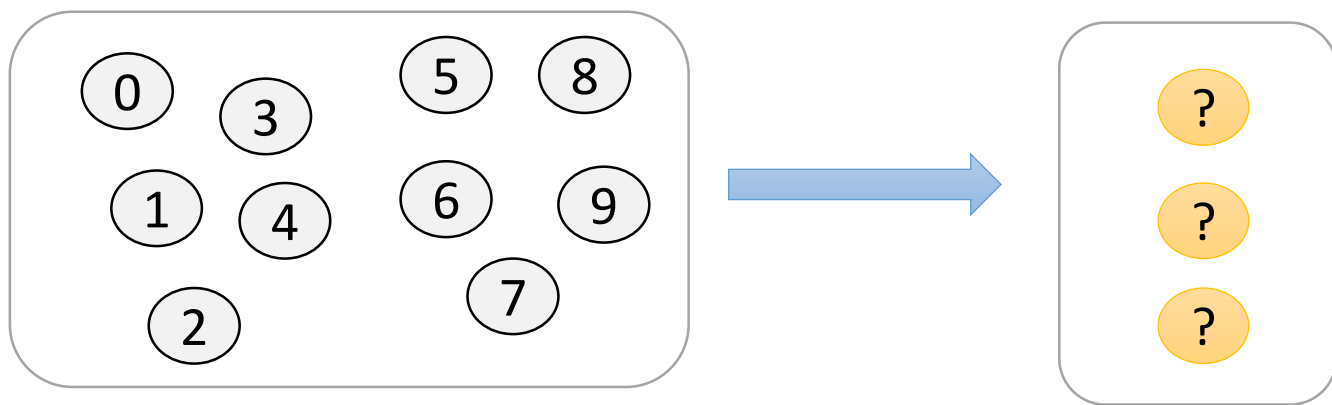
从m个不同元素中取出n($n \leq m$)个元素的所有组合的个数，叫做从m个不同元素中取出n个元素的组合数。记作: $C(m,n)$

$$C(m,n) = C_m^n = \frac{m!}{(m-n)! \cdot n!}$$

案例

在一个盒子中有十个完全相同的球，其中每个球上编有一个编号，球的编号从0到9，求随机抽取3个球，可能出现的数字组合共有多少？

（备注：不考虑数字的顺序，认为1、2、3和3、2、1是一样的）。



解：

$$C_{10}^3 = \frac{10!}{(10-3)!3!} = 120$$

总共10个球，抽取3个球的组合数：

1. 抽取3个球的排列数为 $A(10, 3)$
2. 对于任意排列 (a_1, a_2, a_3) 都有 $3*2*1$ 种相同元素的排列存在
3. 其实组合就是在排列的基础上去掉相同元素后剩下的数量

即：

$$C_{10}^3 = \frac{A_{10}^3}{A_3^3} = \frac{\frac{10!}{(10-3)!}}{\frac{3!}{(3-3)!}} = \frac{10!}{(10-3)!3!} = 120$$

古典概率

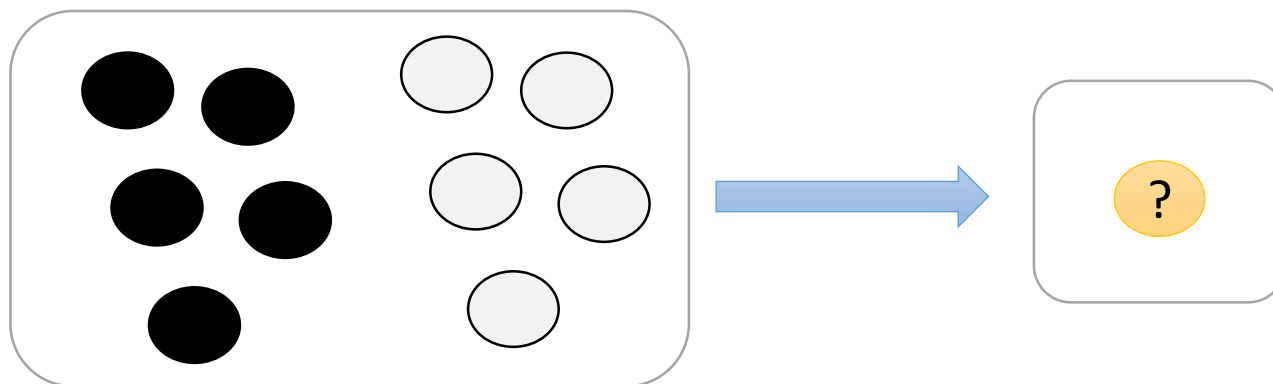
概率是以假设为基础的，即假定随机现象所发生的事件是有限的、互不相容的，而且每个基本事件发生的可能性相等。一般来讲，如果在全部可能出现的基本事件范围内构成事件A的基本事件有a个，不构成事件A的有b个，那么事件A出现的概率为：

$$P(A) = \frac{a}{a+b}$$

概率体现的是随机事件A发生可能的大小度量(数值)

案例

在一个盒子中有十个完全相同的球，其中五个黑球，五个白球，求事件 $A=\{\text{从盒子中获取一个球，颜色是黑色}\}$ 的概率。

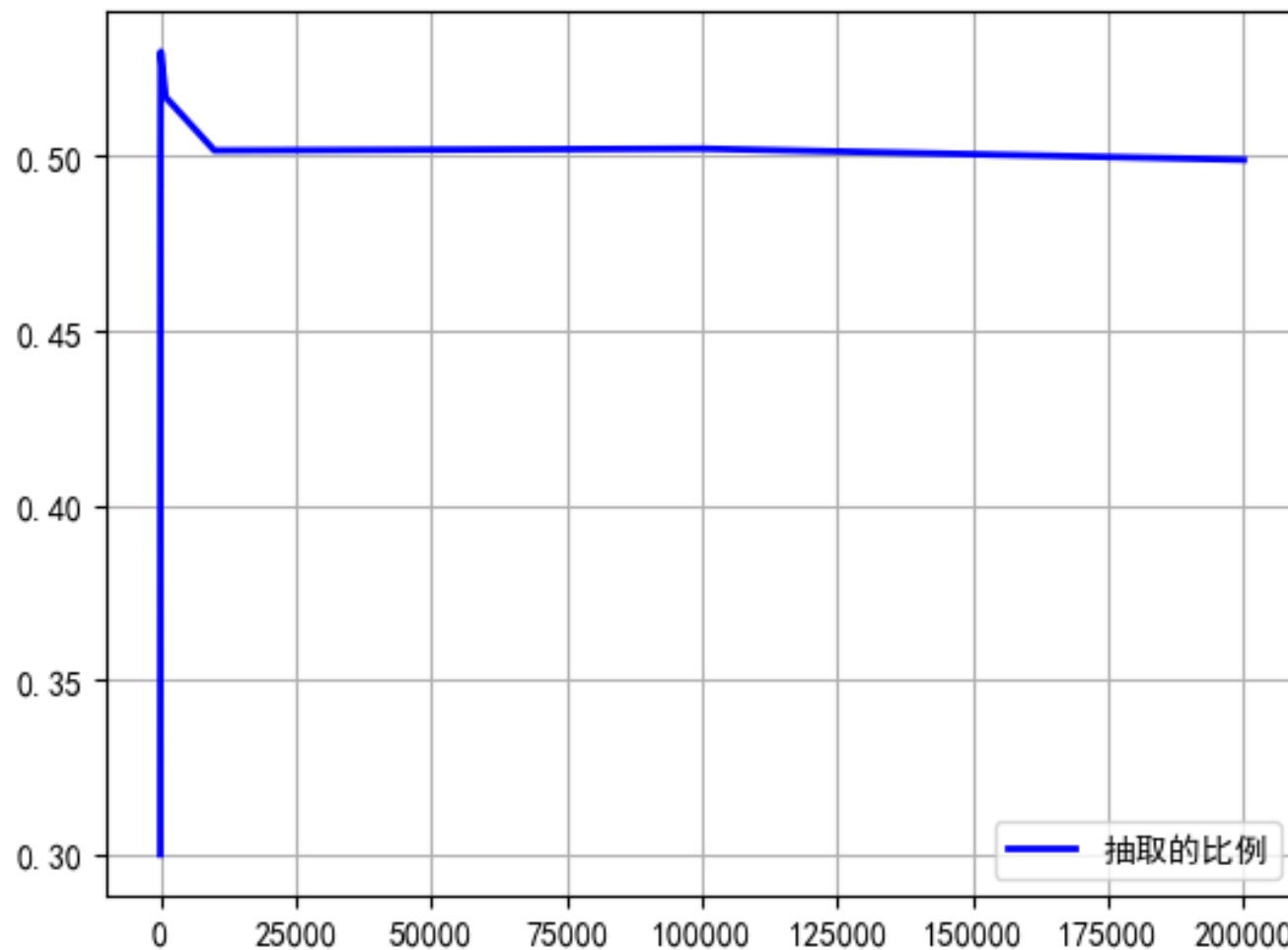


解 $P(A) = \frac{5}{10} = \frac{1}{2}$

基本的事件总数：10

抽取一个球是黑球的事件数：5

[0.3, 0.5, 0.53, 0.517, 0.5017, 0.50223, 0.499045]



例 袋中有 a 只白球， b 只黑球．从中将球取出依次排成一行，问第 k 次取出的球是黑球的 概率．

解： 设 A = “第 k 次取出的球是黑球”

从 $a + b$ 个球中将球取出依次排成一行共有 $(a + b)!$ 种排法（样本点总数）．

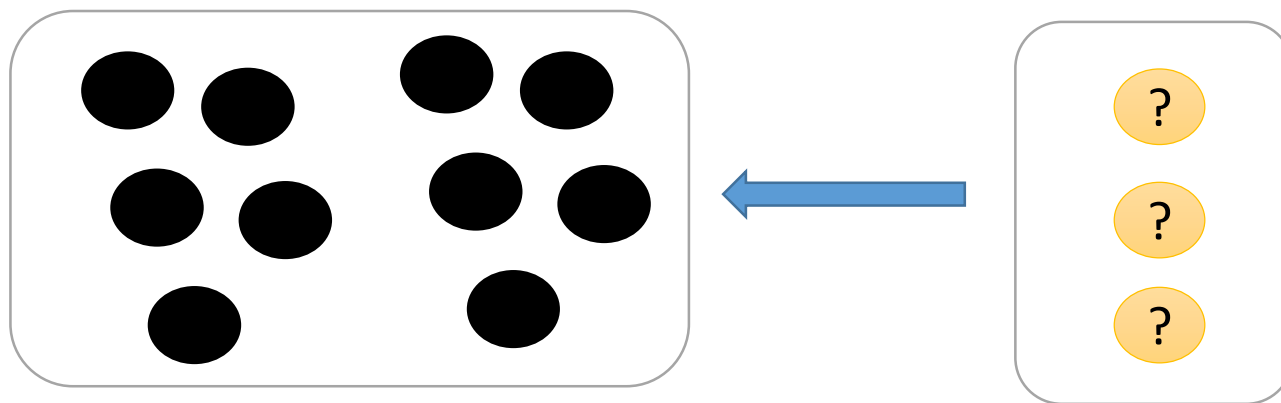
第 k 次取出黑球，有取法 $b(a + b - 1)!$ 种，因此事件 A 所含样本点数为 $b(a + b - 1)!$ ．

所以，

$$P(A) = \frac{b \cdot (a + b - 1)!}{(a + b)!} = \frac{b}{a + b} .$$

案例

假设有 n 个人，每个人都等可能地被分配到 N 个房间中的任意一间去住 ($n \leq N$)，求事件 $A = \{\text{恰好有 } n \text{ 个房间，其中各住一个人}\}$ 的概率



解 $P(A) = \frac{N!}{N^n (N-n)!}$

1. 每个人有N个房间可供选择，所以n个人住的方式共有 N^n 种。
2. 恰好n个房间表示这n个房间其实是从N个房间中任意抽取出来的，也就是从N个房间中抽取n个方法的组合总共有 $C(N,n)$ 种。
3. 对于n个房间来讲，n个人平均分配，那么总共有 $A(n,n)$ 种入住方式。

$$P(A) = \frac{C_N^n A_n^n}{N^n} = \frac{\frac{N!}{(N-n)!} n!}{N^n} = \frac{N!}{N^n (N-n)!}$$

生日问题

北风某个班级有 n 个学生($n \leq 365$)，问至少有两个人的生日在同一天
的概率有多大？

n	10	20	23	30	40	50
P(A)	0.12	0.41	0.51	0.71	0.89	0.97

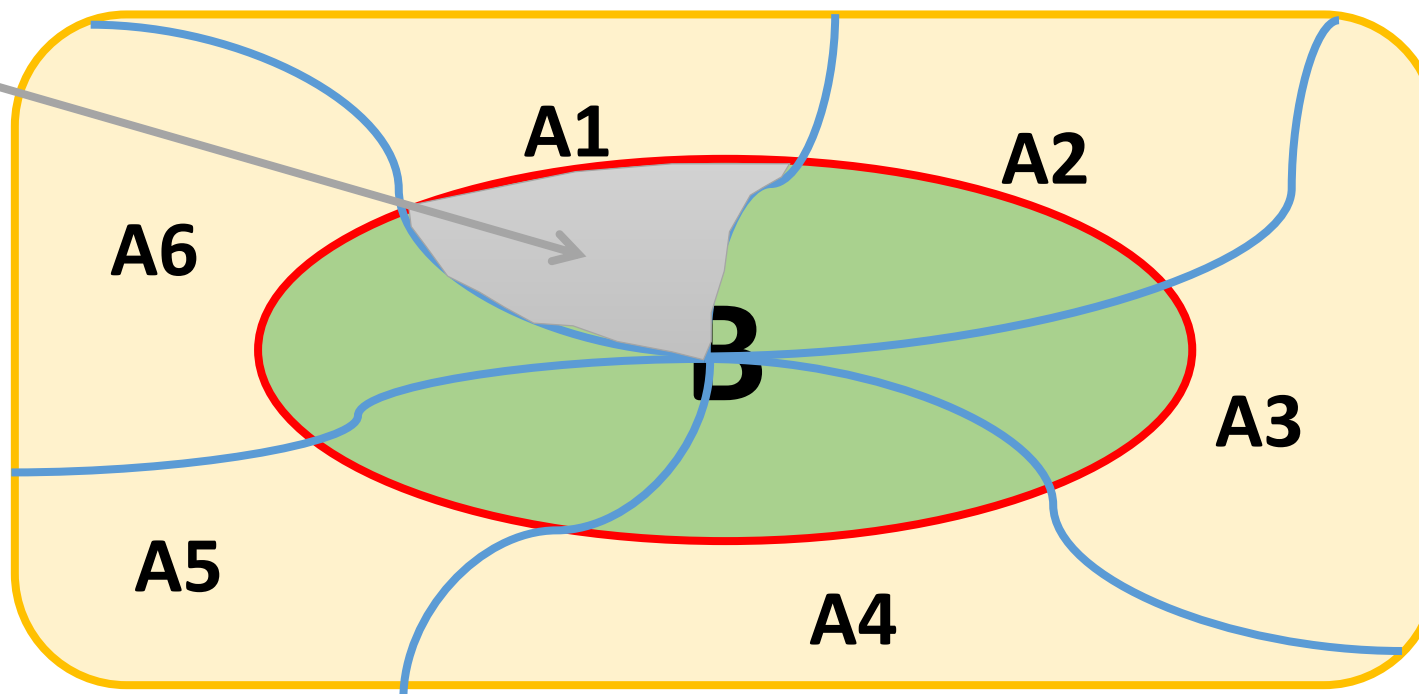
$$P(A) = 1 - \frac{N!}{N^n (N - n)!} \quad N = 365$$

联合概率

表示两个事件共同发生的概率，事件A和事件B的共同概率记作：

$P(AB)$ 、 $P(A,B)$ 或者 $P(A \cap B)$ ，读作“事件A和事件B同时发生的概率”

$P(A_1, B)$



条件概率

事件A在另外一个事件B已经发生的条件下的发生概率叫做条件概率，表示为 $P(A|B)$ ，读作“在B条件下A发生的概率”

一般情况下 $P(A|B) \neq P(A)$ ，而且条件概率具有三个特性：

1. 非负性
2. 可列性
3. 可加性

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

例 1 两台车床加工同一种零件共**100**个，结果如下

	合格品数	次品数	总计
第一台车床加工数	30	5	35
第二台车床加工数	50	15	65
总 计	80	20	100

设 $A=\{ \text{从100个零件中任取一个是合格品} \}$

$B=\{ \text{从100个零件中任取一个是第一台车床加工的} \}$

求： $P(A)$, $P(B)$, $P(AB)$, $P(A|B)$.

解： $P(A) = \frac{80}{100}$, $P(B) = \frac{35}{100}$, $P(AB) = \frac{30}{100}$,

$$P(A|B) = \frac{30}{35} \neq P(A) = \frac{80}{100},$$

例2 已知某家庭有3个小孩，且至少有一个是女孩，求该家庭至少有一个男孩的概率。

解： 设 $A = \{ \text{3个小孩至少有一个女孩} \}$
 $B = \{ \text{3个小孩至少有一个男孩} \}$

所求概率为 $P(B|A) = \frac{P(AB)}{P(A)}$

$$\text{而 } P(A) = 1 - P(\bar{A}) = 1 - \frac{1}{8} = \frac{7}{8} \quad P(AB) = \frac{6}{8}$$

$$\text{所以 } P(B|A) = \frac{\frac{6}{8}}{\frac{7}{8}} = \frac{6}{7}$$

条件概率

将条件概率公式由两个事件推广到任意有穷多个事件时，可以得到如下公式，假设 A_1, A_2, \dots, A_n 为 n 个任意事件($n \geq 2$)，而且 $P(A_1 A_2 \dots A_n) > 0$ ，则：

$$P(A_1 A_2 \dots A_n) = P(A_1) P(A_2 | A_1) \dots P(A_n | A_1 A_2 \dots A_{n-1})$$

例3 袋中有一个白球与一个黑球，现每次从中取出一球，若取出白球，则除把白球放回外再加进一个白球，直至取出黑球为止．求取了 n 次都未取出黑球的概率．

解：

设 $B = \{\text{取了 } n \text{ 次都未取出黑球}\}$

$A_i = \{\text{第 } i \text{ 次取出白球}\} \quad (i = 1, 2, \dots, n)$

则 $B = A_1 A_2 \cdots A_n$

由乘法公式，我们有

$$\begin{aligned} P(B) &= P\left(A_1 A_2 \cdots A_n\right) \\ &= P\left(A_1\right) P\left(A_2 \mid A_1\right) P\left(A_3 \mid A_1 A_2\right) \cdots P\left(A_n \mid A_1 A_2 \cdots A_{n-1}\right) \\ &= \frac{1}{2} \cdot \frac{2}{3} \cdot \frac{3}{4} \cdots \frac{n}{n+1} \\ &= \frac{1}{n+1} \end{aligned}$$

例 4 设某光学仪器厂制造的透镜，第一次落下时打破的概率为 $1/2$ ，若第一次落下未打破，第二次落下打破的概率为 $7/10$ ，若前两次落下未打破，第三次落下打破的概率为 $9/10$ 。求透镜落下三次而未打破的概率。

解：以 A_i ($i=1,2,3$) 表示事件“透镜第 i 次落下打破”，以 B 表示事件“透镜落下三次而未打破”，

有：

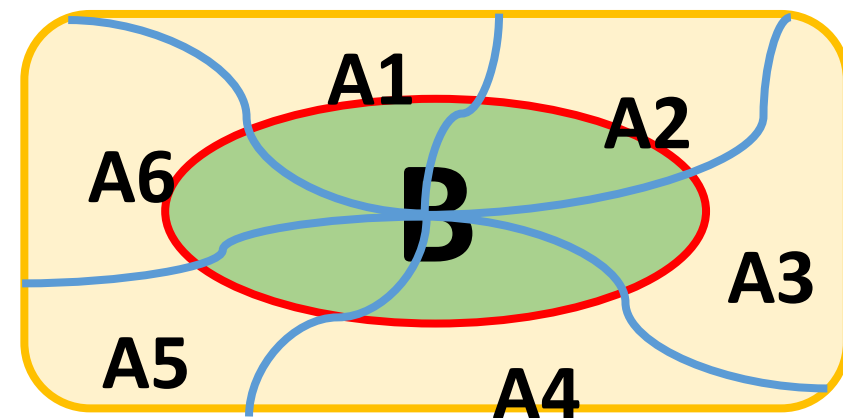
$$\begin{aligned}
 P(B) &= P(\bar{A}_1 \bar{A}_2 \bar{A}_3) \\
 &= P(\bar{A}_1)P(\bar{A}_2 | \bar{A}_1)P(\bar{A}_3 | \bar{A}_1 \bar{A}_2) \\
 &= (1 - \frac{1}{2})(1 - \frac{7}{10})(1 - \frac{9}{10}) = \frac{3}{200} .
 \end{aligned}$$

全概率公式

样本空间 Ω 有一组事件 A_1, A_2, \dots, A_n ，如果事件组满足下列两个条件，那么事件组称为样本空间的一个划分。

$$\forall i \neq j \in \{1, 2, \dots, n\}, A_i A_j = \phi$$

$$A_1 \cup A_2 \cup \dots \cup A_n = \Omega$$



设事件 $\{A_i\}$ 是样本空间 Ω 的一个划分，且 $P(A_i) > 0$ ，

那么对于任意事件 B ，全概率公式为：

$$P(B) = \sum_{i=1}^n P(A_i) P(B | A_i)$$

例5 某小组有20名射手，其中一、二、三、四级射手分别为2、6、9、3名。又若选一、二、三、四级射手参加比赛，则在比赛中射中目标的概率分别为0.85、0.64、0.45、0.32，今随机选一人参加比赛，试求该小组在比赛中射中目标的概率。

解： 设 $B = \{\text{该小组在比赛中射中目标}\}$

$A_i = \{\text{选}i\text{级射手参加比赛}\} (i = 1, 2, 3, 4)$

由全概率公式，有 $P(B) = \sum_{i=1}^4 P(A_i)P(B|A_i)$

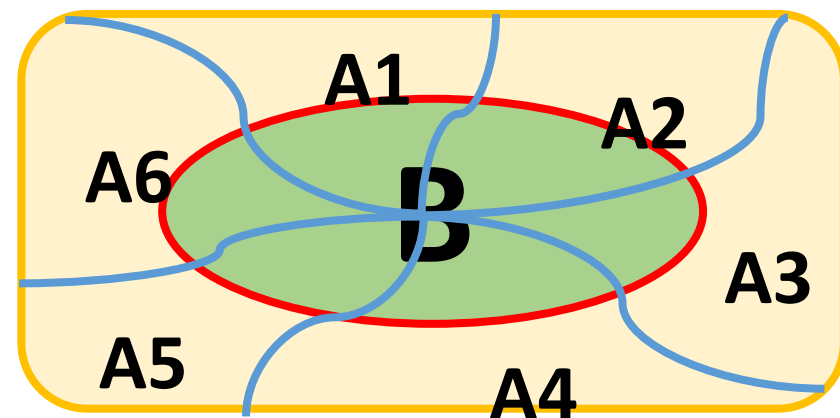
$$\begin{aligned} &= \frac{2}{20} \times 0.85 + \frac{6}{20} \times 0.64 + \frac{9}{20} \times 0.45 + \frac{3}{20} \times 0.32 \\ &= 0.5275 \end{aligned}$$

贝叶斯公式

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

设 A_1, A_2, \dots, A_n 是样本空间 Ω 的一个划分，如果对任意事件 B 而言，有 $P(B) > 0$ ，那么：

$$P(A_i | B) = \frac{P(B, A_i)}{P(B)} = \frac{P(A_i) \cdot P(B | A_i)}{\sum_{j=1}^n P(A_j) \cdot P(B | A_j)}$$



贝叶斯公式案例

一座房子在过去20年里一共发生过2次被盗案，房子的主人养了一条狗，狗平均每周晚上叫3次，在盗贼入侵时狗叫的概率估计为0.9，请求：在狗叫的时候发生入侵的概率是多少？

解： 设 $A = \{ \text{狗叫} \}$ $B = \{ \text{盗贼入侵} \}$

$$P(A) = \frac{3}{7} \quad P(B) = \frac{2}{20 \times 365} = \frac{2}{7300} \quad P(A|B) = 0.9$$

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} = \frac{0.9 \times \frac{2}{7300}}{\frac{3}{7}} = \frac{21}{36500} \approx 0.00058$$

贝叶斯公式

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

$P(A)$ ：在没有数据支持下，A发生的概率：**先验概率**或**边缘概率**。

$P(A|B)$ ：在已知B发生后A的条件概率，也就是由于得自B的取值而被称为A的**后验概率**。

$P(B|A)$ ：在已知A发生的情况下的概率分布：似然函数。

例 6 用某种方法普查肝癌，设：

$A = \{ \text{用此方法判断被检查者患有肝癌} \}$,

$D = \{ \text{被检查者确实患有肝癌} \}$,

已知

$$P(A|D) = 0.95, \quad P(\bar{A}|\bar{D}) = 0.90$$

而且已知： $P(D) = 0.0004$

现有一人用此法检验患有肝癌，求此人真正患有肝癌的概率。

解： 由已知，得 $P(A|D)=0.95$, $P(\bar{A}|\bar{D})=0.90$
 $P(D)=0.0004$

所以，由Bayes公式，得

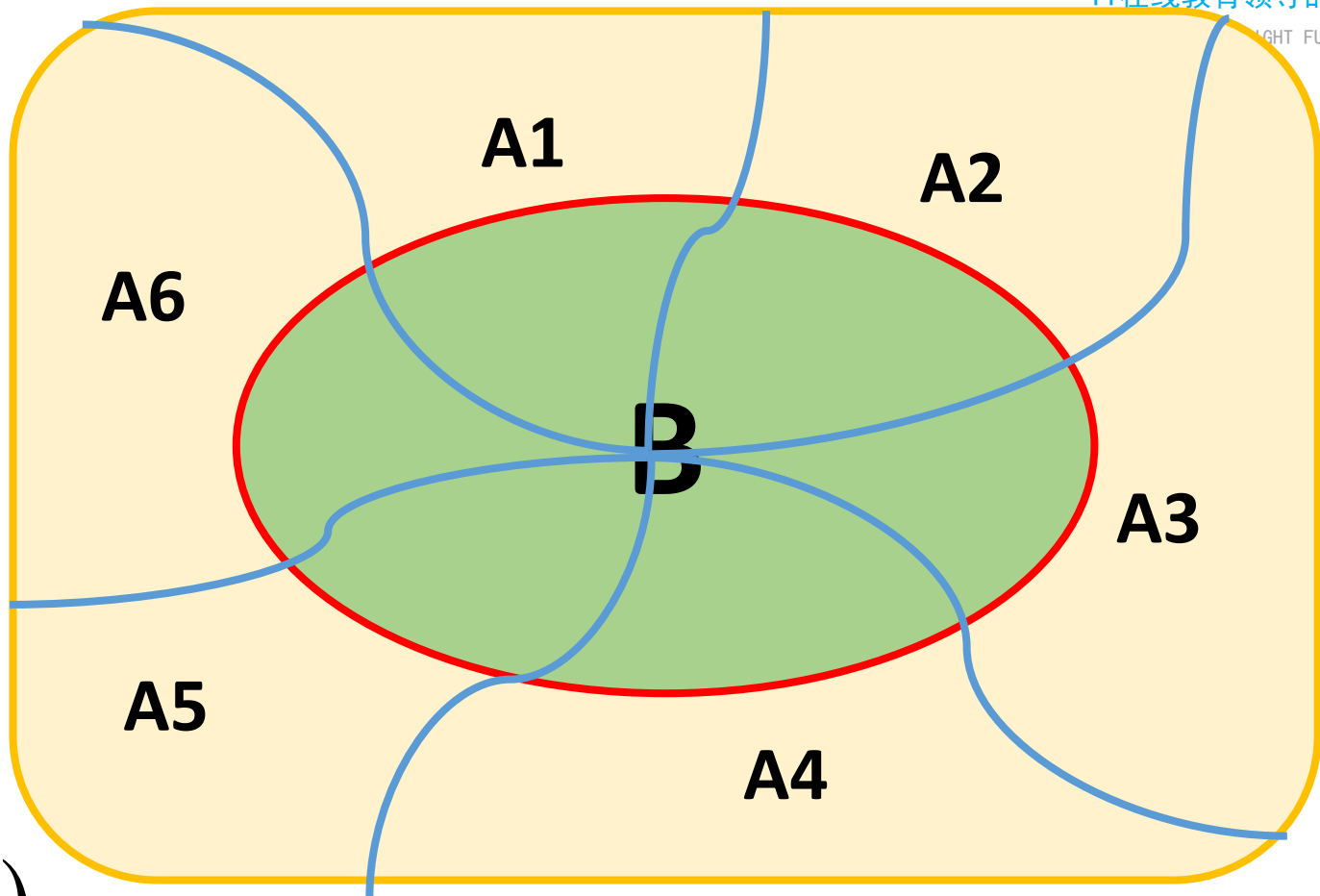
$$\begin{aligned} P(D|A) &= \frac{P(DA)}{P(A)} = \frac{P(D)P(A|D)}{P(D)P(A|D) + P(\bar{D})P(A|\bar{D})} \\ &= \frac{0.0004 \times 0.95}{0.0004 \times 0.95 + 0.9996 \times 0.10} \\ &= 0.0038 \end{aligned}$$

概率公式

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

$$P(B) = \sum_{i=1}^n P(A_i)P(B | A_i)$$

$$P(A_i | B) = \frac{P(B | A_i)P(A_i)}{\sum_{j=1}^n P(A_j)P(B | A_j)}$$



事件的独立性

给定A、B两个事件，如果概率存在 $P(A,B)=P(A)P(B)$ （A、B事件的联合分布概率等于事件各自分布概率的积），则事件A和B相互独立。

如果事件A、B相互独立，互不影响，那么存在 $P(A|B)=P(A)$ ，
 $P(B|A)=P(B)$

例1 袋中装有 4 个外形相同的球，其中三个球分别涂有红、白、黑色，另一个球涂有红、白、黑三种颜色．现从袋中任意取出一球，令：

$A = \{ \text{取出的球涂有红色} \}$

$B = \{ \text{取出的球涂有白色} \}$

$C = \{ \text{取出的球涂有黑色} \}$

则：

$$\begin{aligned} P(A) &= P(B) = P(C) = \frac{1}{2} \\ P(AB) &= P(BC) = P(AC) = \frac{1}{4} \\ P(ABC) &= \frac{1}{4} \end{aligned}$$

由此可见 $P(AB) = P(A)P(B)$,

$$P(BC) = P(B)P(C), \quad P(AC) = P(A)P(C).$$

但是
$$P(ABC) = \frac{1}{4} \neq \frac{1}{8} = P(A)P(B)P(C)$$

这表明， A 、 B 、 C 这三个事件是两两独立的，但不是相互独立的。

随机变量及其分布

随机变量

例 1 袋中有3只黑球，2只白球，从中任意取出3只球。我们将3只黑球分别记作1，2，3号，2只白球分别记作4，5号，则该试验的样本空间为

$$S = \left\{ \begin{array}{lll} (1, 2, 3) & (1, 2, 4) & (1, 2, 5) \\ (1, 3, 4) & (1, 3, 5) & (1, 4, 5) \\ (2, 3, 4) & (2, 3, 5) & (2, 4, 5) \\ (3, 4, 5) \end{array} \right\}$$

考察取出的3只球中的黑球的个数。

我们记取出的**黑球数为 X** ，则 **X** 的可能取值为**1, 2, 3**。因此， X 是一个变量。但是， X 取什么值依赖于试验结果，即 X 的取值带有随机性，所以，我们称 **X 为随机变量**。 X 的取值情况可由下表给出：

样本点	黑球数 X	样本点	黑球数 X
(1, 2, 3)	3	(1, 4, 5)	1
(1, 2, 4)	2	(2, 3, 4)	2
(1, 2, 5)	2	(2, 3, 5)	2
(1, 3, 4)	2	(2, 4, 5)	1
(1, 3, 5)	2	(3, 4, 5)	1

由上表可以看出，该随机试验的每一个结果都对应着变量 X 的一个确定的取值，因此变量 X 是样本空间 S 上的函数：

$$X = X(e) \quad (e \in S)$$

我们定义了随机变量后，就可以用随机变量的取值情况来刻画随机事件。例如

$$\{e: X(e) = 2\} = \{X = 2\}$$

表示取出2个黑球这一事件；

$$\{X \geq 2\}$$

表示至少取出2个黑球这一事件，等等。

通常随机变量用大写的英文字母 X 、 Y 、 Z 、...

或希腊字母 ξ 、 η 、 ς 、...等来表示.

例2 掷一颗骰子, 令 X : 出现的点数.

则 X 就是一个随机变量. 它的取值为1, 2, 3, 4, 5, 6.

$$\{X \leq 4\}$$

表示掷出的点数不超过 4 这一随机事件;

$$\{X \text{ 取偶数}\}$$

表示掷出的点数为偶数这一随机事件.

例3 上午 8:00~9:00 在某路口观察, 令:

Y : 该时间间隔内通过的汽车数.

则 Y 就是一个随机变量. 它的取值为 $0, 1, \dots$

$$\{Y < 100\}$$

表示通过的汽车数小于100辆这一随机事件;

$$\{50 < Y \leq 100\}$$

表示通过的汽车数大于 50 辆但不超过 100 辆这一随机事件.

注意 Y 的取值是可列无穷个!

例 4 观察某电子元件的寿命（单位：小时），令
 Z ：该电子元件的寿命。

则 Z 就是一个随机变量。它的取值为所有非负实数。

$$\{Z \leq 500\}$$

表示该电子元件的寿命不超过500小时这一随机事件。

$$\{Z > 1000\}$$

表示该电子元件的寿命大于 1000小时这一随机事件。

注意 Z 的取值是不可列无穷个！

例 5 掷一枚硬币，令：

$$X = \begin{cases} 1 & \text{掷硬币出现正面} \\ 0 & \text{掷硬币出现反面} \end{cases}$$

则 X 是一个随机变量.

说 明：

在同一个样本空间上可以定义不同的随机变量.

例 6 掷一枚骰子，在[例2](#)中，我们定义了随机变量 X 表示出现的点数。我们还可以定义其它的随机变量，例如我们可以定义：

$$Y = \begin{cases} 1 & \text{出现偶数点;} \\ 0 & \text{出现奇数点.} \end{cases}$$

$$Z = \begin{cases} 1 & \text{点数为6;} \\ 0 & \text{点数不为6.} \end{cases}$$

等等.

离散型随机变量及其分布律

一、离散型随机变量的分布律与性质

1) 离散型随机变量的定义

如果随机变量 X 的取值是有限个或可列无穷个，则称 X 为离散型随机变量。

2)离散型随机变量的分布律

设离散型随机变量 X 的所有可能取值为

$$x_1, x_2, \dots, x_n, \dots$$

并设 $P\{X = x_n\} = p_n \quad (n = 1, 2, \dots)$

则称上式或

X	x_1	x_2	$\dots,$	x_n	\dots
P	p_1	p_2	$\dots,$	p_n	\dots

为离散型随机变量 X 的分布律.

3) 离散型随机变量分布律的性质:

(1) 对任意的自然数 n , 有 $p_n \geq 0$;

$$(2) \sum_n p_n = 1.$$

例 1 从1~10这10个数字中随机取出5个数字，令 X ：取出的5个数字中的最大值。试求 X 的分布律。

解： X 的可能取值为 5, 6, 7,

求分布律一定要说明 k 的取值范围！

$$P\{X = k\} = \frac{C_{k-1}^4}{C_{10}^5} \quad k = 5, 6, \dots, 10.$$

具体写出，即可得 X 的分布律：

X	5	6	7	8	9	10
P	$\frac{1}{252}$	$\frac{5}{252}$	$\frac{15}{252}$	$\frac{35}{252}$	$\frac{70}{252}$	$\frac{126}{252}$

例 2 将 1 枚硬币掷 3 次，令

X : 出现的正面次数与反面次数之差.

试求: (1) X 的分布律; (2) $P\{0.5 \leq X < 3\}$.

解: X 的可能取值为 -3, -1, 1, 3. 并且分布率为

X	-3	-1	1	3
P_k	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

$$P\{0.5 \leq X < 3\} = P\{X = 1\} = \frac{3}{8}.$$

例 3 设随机变量 X 的分布律为

$$P\{X = n\} = c \left(\frac{1}{4}\right)^n \quad (n = 1, 2, \dots) \quad \text{试求常数 } c.$$

解： 由分布率的性质，得

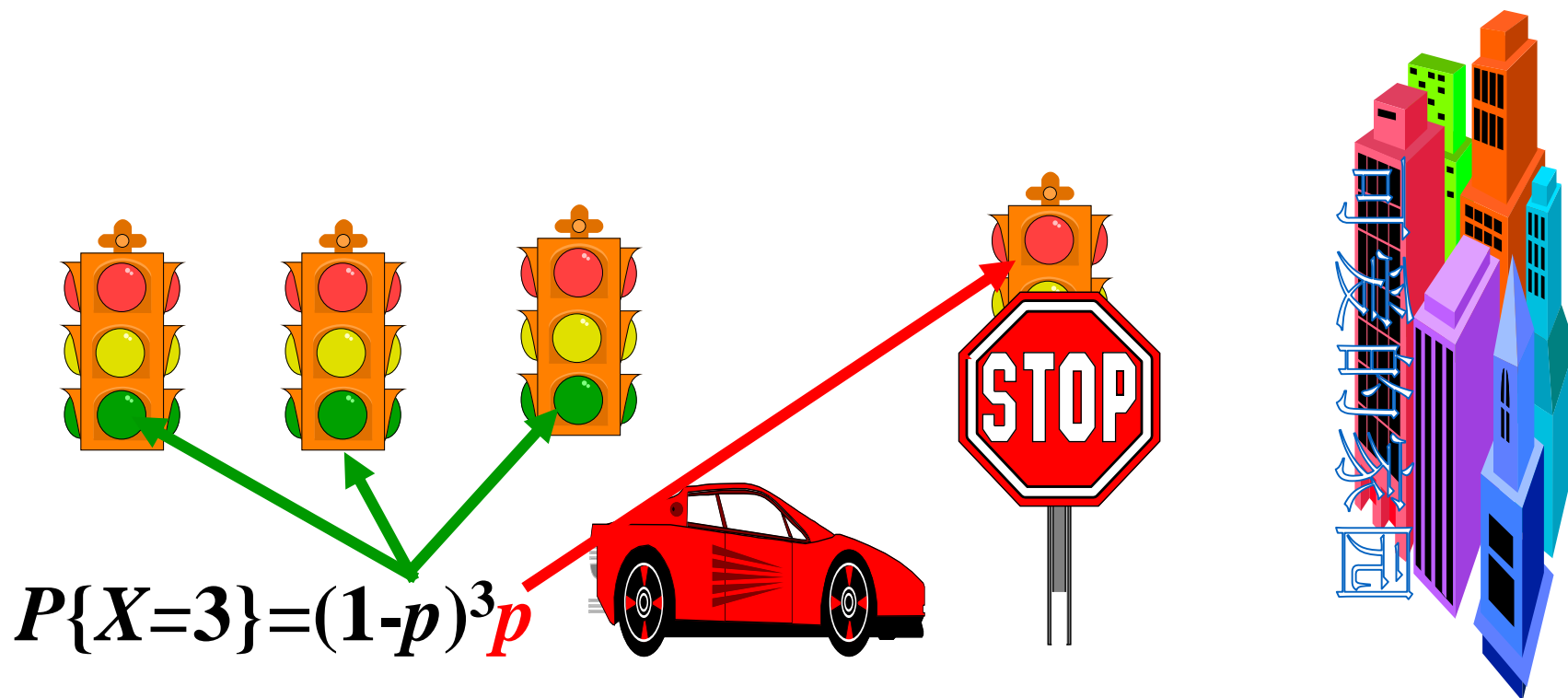
$$1 = \sum_{n=1}^{\infty} P\{X = n\} = \sum_{n=1}^{\infty} c \left(\frac{1}{4}\right)^n$$

该级数为等比级数，故有

$$1 = \sum_{n=1}^{\infty} c \left(\frac{1}{4}\right)^n = c \cdot \frac{\frac{1}{4}}{1 - \frac{1}{4}} = \frac{c}{3}$$

所以 $c = 3$.

例 4 设一汽车在开往目的地的道路上需经过四盏信号灯，每盏信号灯以概率 p 禁止汽车通过. 以 X 表示汽车首次停下时，它已通过的信号灯的盏数，求 X 的分布律. (信号灯的工作是相互独立的).



解： 以 p 表示每盏信号灯禁止汽车通过的概率，则 X 的分布律为：

X	0	1	2	3	4
p_k	p	$(1-p)p$	$(1-p)^2p$	$(1-p)^3p$	$(1-p)^4$

或写成 $P\{X=k\} = (1-p)^k p, k = 0, 1, 2, 3$

$$P\{X=4\} = (1-p)^4$$

以 $p = 1/2$ 代入得：

X	0	1	2	3	4
p_k	0.5	0.25	0.125	0.0625	0.0625

二、一些常用的离散型随机变量

1) *Bernoulli*分布

如果随机变量 X 的分布律为

$$P\{X = k\} = p^k (1-p)^{1-k}, k = 0, 1$$

或

X	0	1
P	$1-p$	p

则称随机变量 X 服从参数为 p 的 *Bernoulli*分布.

记作 $X \sim B(1, p)$ (其中 $0 \leq p \leq 1$ 为参数)

*Bernoulli*分布也称作 **0-1 分布**或**二点分布**.

*Bernoulli*分布的概率背景

进行一次*Bernoulli*试验， A 是随机事件。设：

$$P(A) = p, \quad P(\bar{A}) = 1 - p = q$$

设 X 表示这次*Bernoulli*试验中事件 A 发生的次数.
或者设

$$X = \begin{cases} 1 & \text{若事件} A \text{发生} \\ 0 & \text{若事件} A \text{不发生} \end{cases}$$

则 $X \sim B(1, p)$

2) 二 项 分 布

如果随机变量 X 的分布律为

$$P\{X = k\} = C_n^k p^k (1-p)^{n-k} \quad (k = 0, 1, \dots, n)$$

则称随机变量 X 服从参数为 (n, p) 的二项分布。

记作 $X \sim B(n, p)$

(其中 n 为自然数, $0 \leq p \leq 1$ 为参数)



说 明

显然，当 $n=1$ 时

$$X \sim B(1, p)$$

此时， X 服从 *Bernoulli* 分布。

这说明，*Bernoulli* 分布是二项分布的一个特例。

二项分布的概率背景

进行 n 重 *Bernoulli* 试验， A 是随机事件。设在每次试验中

$$P(A) = p, \quad P(\bar{A}) = 1 - p = q$$

令 X 表示这 n 次 *Bernoulli* 试验中事件 A 发生的次数.

则 $X \sim B(n, p)$

说明：

设 $A_i = \{\text{第 } i \text{ 次试验 } A \text{ 出现}\}$, 则

$$\begin{aligned} \{X = k\} = & A_1 \cdots A_k \bar{A}_{k+1} \cdots \bar{A}_n \cup \bar{A}_1 A_2 \cdots A_{k+1} \bar{A}_{k+2} \cdots \bar{A}_n \cup \\ & \cup \cdots \cup \bar{A}_1 \bar{A}_2 \cdots \bar{A}_{n-k} A_{n-k+1} \cdots A_n \end{aligned}$$

在 n 次试验中，指定 k 次出现 A (成功)，其余 $n - k$ 次出现 \bar{A} (失败)，这种指定的方法共有 C_n^k 种。

所以

$$\begin{aligned} P\{X = k\} = & C_n^k p^k q^{n-k} \quad (q = 1 - p) \\ & (k = 0, 1, 2, \cdots, n) \end{aligned}$$

例5 一大批产品的次品率为0.1，现从中取出15件．试求下列事件的概率：

$B = \{ \text{取出的15件产品中恰有2件次品} \}$

$C = \{ \text{取出的15件产品中至少有2件次品} \}$

解：由于从一大批产品中取15件产品，故可近似看作是一15重*Bernoulli*试验．

$A = \{ \text{取出一件产品为次品} \}$ ，则 $P(A) = 0.1$ ．

所以， $P(B) = C_{15}^2 \times 0.1^2 \times 0.9^{13}$

$$\begin{aligned} P(C) &= 1 - P(\bar{C}) \\ &= 1 - C_{15}^0 \times 0.1^0 \times 0.9^{15} - C_{15}^1 \times 0.1 \times 0.9^{14} \end{aligned}$$

例 6 一张考卷上有5道选择题，每道题列出4个可能答案，其中只有一个答案是正确的．某学生靠猜测能答对4道题以上的概率是多少？

解： 每答一道题相当于做一次*Bernoulli*试验，

$$A = \{\text{答对一道题}\}, \quad \text{则} \quad P(A) = \frac{1}{4}$$

则答5道题相当于做5重*Bernoulli*试验．

设 X 表示该学生靠猜测能答对的题数，

$$\text{则 } X \sim B\left(5, \frac{1}{4}\right)$$

所以

$$\begin{aligned}P\{\text{至少能答对4道题}\} &= P\{X \geq 4\} \\&= P\{X = 4\} + P\{X = 5\} \\&= C_5^4 \left(\frac{1}{4}\right)^4 \cdot \frac{3}{4} + \left(\frac{1}{4}\right)^5 \\&= \frac{1}{64}\end{aligned}$$

二项分布的分布形态

若 $X \sim B(n, p)$, 则

$$\frac{P\{X = k\}}{P\{X = k - 1\}} = 1 + \frac{(n + 1)p - k}{kq} \quad (q = 1 - p)$$

由此可知, 二项分布的分布率

$$P\{X = k\}$$

先是随着 k 的增大而增大, 达到其最大值后再随着 k 的增大而减少. 这个使得

$$P\{X = k\}$$

达到其最大值的 k_0 称为该二项分布的最可能次数.

$$\frac{P\{X = k\}}{P\{X = k - 1\}} = 1 + \frac{(n + 1)p - k}{kq} \quad (q = 1 - p)$$

可以证明：

如果 $(n + 1)p$ 不是整数，则 $k_0 = [(n + 1)p]$ ；

如果 $(n + 1)p$ 是整数，则 $k_0 = (n + 1)p$ 或 $(n + 1)p - 1$ ；

例 7 对同一目标进行300次独立射击，设每次射击时的命中率均为0.44，试求300次射击最可能命中几次？其相应的概率是多少？

解：对目标进行300次射击相当于做300重*Bernoulli*试验．令：

X 表示300射击中命中目标的次数

则由题意 $X \sim B(300, 0.44)$.

由于 $(300+1) \times 0.44 = 132.44$ ，它不是整数

因此，最可能射击的命

$$\text{中次数为 } k_0 = [132.44] = 132$$

其相应的概率为

$$\begin{aligned} P\{X = 132\} &= C_{300}^{132} \times 0.44^{132} \times 0.56^{168} \\ &= 0.04636 \end{aligned}$$

3) *Poisson* 分布

如果随机变量 X 的分布律为

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots)$$

(其中 $\lambda > 0$ 为常数)

则称随机变量 X 服从参数为 λ 的 *Poisson* 分布.

Poisson 分布的应用

- *Poisson*分布是概率论中重要的分布之一.
- 自然界及工程技术中的许多随机指标都服从*Poisson*分布.
- 例如, 可以证明, 电话总机在某一时间间隔内收到的呼叫次数, 放射物在某一时间间隔内发射的粒子数, 容器在某一时间间隔内产生的细菌数, 某一时间间隔内来到某服务台要求服务的人数, 等等, 在一定条件下, 都是服从*Poisson*分布的.

例11 如果随机变量 X 的分布律为

$$P\{X = k\} = c \frac{\lambda^k}{k!} \quad (k = 1, 2, \dots).$$

(其中 $\lambda > 0$ 为常数) 试确定未知常数 c .

解: 由分布率的性质有 $\sum_{k=1}^{\infty} c \frac{\lambda^k}{k!} = c \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} = 1,$

$$\text{而 } \sum_{k=1}^{\infty} \frac{\lambda^k}{k!} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} - 1 = e^{\lambda} - 1$$

$$\text{所以 } c = \frac{1}{e^{\lambda} - 1}.$$

例 12 设随机变量 X 服从参数为 λ 的 *Poisson* 分布, 且已知

$$P\{X = 1\} = P\{X = 2\}$$

试求 $P\{X = 4\}$.

解: 随机变量 X 的分布律为

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda} \quad (k = 0, 1, 2, \dots)$$

由已知 $P\{X = 1\} = P\{X = 2\}$

得

$$\frac{\lambda^1}{1!} e^{-\lambda} = \frac{\lambda^2}{2!} e^{-\lambda}$$

由此得方程 $\lambda^2 - 2\lambda = 0$

得解 $\lambda = 2.$

(另一个解 $\lambda = 0$ 不合题意, 舍去)

所以,

$$P\{X = 4\} = \frac{2^4}{4!} e^{-2} = \frac{2}{3} e^{-2}$$

$$= 0.09022$$

Poisson 定理

设在 *Bernoulli* 试验中，以 p_n 代表事件 A 在试验中发生的概率，它与试验总数 n 有关．如果

$$\lim_{n \rightarrow \infty} np_n = \lambda > 0$$

则
$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

证明： 令： $np_n = \lambda_n$

$$\begin{aligned} & \text{则} \quad C_n^k p_n^k (1 - p_n)^{n-k} \\ &= \frac{n(n-1)(n-2) \cdots (n-k+1)}{k!} \left(\frac{\lambda_n}{n} \right)^k \left(1 - \frac{\lambda_n}{n} \right)^{n-k} \end{aligned}$$

$$= \frac{\lambda^k}{k!} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

对于固定的 k , 有 $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} = e^{-\lambda}$

所以,

$$\begin{aligned} & \lim_{n \rightarrow \infty} C_n^k p_n^k (1 - p_n)^{n-k} \\ &= \frac{1}{k!} \lim_{n \rightarrow \infty} \lambda^k \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \cdots \left(1 - \frac{k-1}{n}\right) \cdot \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{\lambda^k}{k!} e^{-\lambda} \end{aligned}$$

Poisson定理的应用

由 *Poisson* 定理, 可知

若随机变量 $X \sim B(n, p)$,

则当 n 比较大, p 比较小时,

令: $\lambda = np$

则有 $P\{X = k\} = C_n^k p^k (1-p)^{n-k}$

$$\approx \frac{\lambda^k}{k!} e^{-\lambda}$$

例13 设每次射击命中目标的概率为0.012，现射击600次，求至少命中3次目标的概率（用*Poisson*分布近似计算）。

解：设 X 为600次射击命中目标的次数

则 $X \sim B(600, 0.012)$. 取 $\lambda = 600 \times 0.012 = 7.2$.

$$\begin{aligned} P\{X \geq 3\} &= 1 - P\{X < 3\} \\ &= 1 - P\{X = 0\} - P\{X = 1\} - P\{X = 2\} \\ &= 1 - e^{-7.2} - 7.2e^{-7.2} - \frac{7.2^2}{2}e^{-7.2} \\ &= 0.9745 \end{aligned}$$

例 14 保险公司售出某种寿险（一年）保单2500份.每单交保费100元，当被保险人一年内死亡时，家属可从保险公司获得2万元的赔偿.若此类被保险人一年内死亡的概率为0.001，求

- (1) 保险公司亏本的概率；
- (2) 保险公司获利不少于10万元的概率.

解： 设此类被保险人一年内死亡的人数为 X ，
则 $X \sim b(2500, 0.001)$.

$$(1) P(\text{保险公司亏本}) = P(25 - 2X < 0)$$

$$= 1 - P(X \leq 12) = 1 - \sum_{k=0}^{12} \binom{2500}{k} \cdot (0.001)^k \cdot (0.999)^{2500-k}$$

$$\approx 1 - \sum_{k=0}^{12} \frac{2.5^k}{k!} e^{-2.5} = 0.000002$$

$$(2) P(\text{保险公司获利不少于10万元}) = P(25 - 2X \geq 10)$$

$$= P(X \leq 7) \approx \sum_{k=0}^7 \frac{2.5^k}{k!} e^{-2.5} = 0.995753$$

4) 几何分布

若随机变量 X 的分布律为

$$P\{X = k\} = q^{k-1} p \quad (k = 1, 2, \dots)$$

$$(其中\ p \geq 0, \ q \geq 0, \ p + q = 1)$$

则称随机变量 X 服从参数为 p 的几何分布.

几何分布的概率背景

在*Bernoulli*试验中,

$$P(A) = p, \quad P(\bar{A}) = q = 1 - p$$

试验进行到 A 首次出现为止.

令: X 表示所需试验次数

则 X 服从参数为 p 的几何分布.

$$\text{即 } P\{X = k\} = q^{k-1} p \quad (k = 1, 2, \dots)$$

5) 超几何分布

如果随机变量 X 的分布律为

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k = 0, 1, \dots, \min(M, n))$$

其中 N, M, n 均为自然数.

则称随机变量 X 服从参数为 (N, M, n) 的超几何分布.

超几何分布的概率背景

一批产品有 N 件，其中有 M 件次品，其余 $N-M$ 件为正品．现从中取出 n 件．

令 X ：取出 n 件产品中的次品数． 则 X 的分布律为

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n} \quad (k = 0, 1, \dots, \min(M, n))$$

此时，随机变量 X 服从参数为 (N, M, n) 的超几何分布

连续型随机变量及其概率密度

一、连续型随机变量的概念与性质

1) 定义 如果对于随机变量 X 的分布函数 $F(x)$, 存在非负函数 $f(x)$, 使得对于任意实数 x , 有

$$F(x) = \int_{-\infty}^x f(t)dt,$$

则称 X 为连续型随机变量, 其中函数 $f(x)$ 称为 X 的概率密度函数, 简称概率密度.

六、不定积分(以下为补充内容)

定义 1. $f(x)$ 在区间 I 上的原函数全体称为 $f(x)$ 在 I 上的不定积分, 记作 $\int f(x)dx$, 其中

\int — 积分号; $f(x)$ — 被积函数;
 x — 积分变量; $f(x)dx$ — 被积表达式.

若 $F'(x) = f(x)$, 则

$$\int f(x) dx = F(x) + C \quad (C \text{ 为任意常数})$$

例如, $\int e^x dx = e^x + C$

$$\int x^2 dx = \frac{1}{3} x^3 + C$$

$$\int \sin x dx = -\cos x + C$$

C 称为积分常数
不可丢!

不定积分的性质

性质 1 设函数 $f(x)$ 及 $g(x)$ 的原函数存在，则

$$\int [f(x) \pm g(x)] dx = \int f(x) dx \pm \int g(x) dx$$

性质 2 设函数 $f(x)$ 原函数存在， k 为非零常数，则

$$\int k f(x) dx = k \int f(x) dx \quad (k \neq 0)$$

推论：若 $f(x) = \sum_{i=1}^n k_i f_i(x)$ ，则

$$\int f(x) dx = \sum_{i=1}^n k_i \int f_i(x) dx$$

从不定积分定义可知:

$$(1) \quad \frac{d}{dx} \left[\int f(x) dx \right] = f(x) \quad \text{或} \quad d \left[\int f(x) dx \right] = f(x) dx$$

$$(2) \quad \int F'(x) dx = F(x) + C \quad \text{或} \quad \int dF(x) = F(x) + C$$

基本积分表

利用逆向思维

$$(1) \quad \int k dx = kx + C \quad (k \text{ 为常数})$$

$$(2) \quad \int x^{\mu} dx = \frac{1}{\mu+1} x^{\mu+1} + C \quad (\mu \neq -1)$$

$$(3) \quad \int \frac{dx}{x} = \ln|x| + C$$

$$(\ln|x|)' = \frac{1}{x}$$

$$(4) \quad \int \frac{dx}{1+x^2} = \arctan x + C \text{ 或 } -\operatorname{arccot} x + C$$

$$(5) \quad \int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C \text{ 或 } -\arccos x + C$$

$$(6) \quad \int \cos x dx = \sin x + C$$

$$(7) \quad \int \sin x dx = -\cos x + C$$

$$(8) \quad \int \frac{dx}{\cos^2 x} = \int \sec^2 x dx = \tan x + C$$

$$(9) \quad \int \frac{dx}{\sin^2 x} = \int \csc^2 x dx = -\cot x + C$$

$$(10) \quad \int \sec x \tan x dx = \sec x + C$$

$$(11) \quad \int \csc x \cot x dx = -\csc x + C$$

$$(12) \quad \int e^x dx = e^x + C$$

$$(13) \quad \int a^x dx = \frac{a^x}{\ln a} + C$$

$$(14) \quad \int \frac{1}{x^2} dx = -\frac{1}{x} + C$$

$$(15) \quad \int \frac{1}{\sqrt{x}} dx = 2\sqrt{x} + C$$

例1. 求 $\int \frac{dx}{x \sqrt[3]{x}} \cdot$

解: 原式 = $\int x^{-\frac{4}{3}} dx$

$$= \frac{x^{-\frac{4}{3}+1}}{-\frac{4}{3}+1} + C$$

$$= -3x^{-\frac{1}{3}} + C$$

$$(2) \quad \int x^{\mu} dx = \frac{1}{\mu+1} x^{\mu+1} + C \quad (\mu \neq -1)$$

例2. 求 $\int 2^x (e^x - 5) dx$.

$$\begin{aligned} \text{解: 原式} &= \int [(2e)^x - 5 \cdot 2^x] dx \\ &= \int (2e)^x dx - 5 \int 2^x dx \\ &= \frac{(2e)^x}{\ln(2e)} - 5 \frac{2^x}{\ln 2} + C \\ &= 2^x \left[\frac{e^x}{\ln 2 + 1} - \frac{5}{\ln 2} \right] + C \end{aligned}$$

$$(13) \quad \int a^x dx = \frac{a^x}{\ln a} + C$$

$[a, b]$ 称为积分区间

积分上限

积分下限

$$\int_a^b f(x) dx = \lim_{\lambda \rightarrow 0} \sum_{i=1}^n f(\xi_i) \Delta x_i$$

被积函数

被积表达式

积分变量

积分和

注意：积分变量的变化区间是积分区间。

引例1. 曲边梯形面积

$$A = \int_a^b f(x) dx \quad (f(x) \geq 0)$$

引例2. 变速直线运动的路程

$$s = \int_{T_1}^{T_2} v(t) dt$$

牛顿-莱布尼茨公式

定理 设 $F(x)$ 是连续函数 $f(x)$ 在 $[a, b]$ 上的一个原函数，则 $\int_a^b f(x) dx = F(b) - F(a)$. (牛顿-莱布尼茨公式)

例1 求 $\int_{-2}^{-1} \frac{1}{x} dx$.

解 当 $x < 0$ 时， $\frac{1}{x}$ 的一个原函数是 $\ln |x|$,

$$\int_{-2}^{-1} \frac{1}{x} dx = [\ln |x|]_{-2}^{-1} = \ln 1 - \ln 2 = -\ln 2.$$

例2 求 $\int_0^{\frac{\pi}{2}} (2\cos x + \sin x - 1) dx$.

解 原式 $= [2\sin x - \cos x - x]_0^{\frac{\pi}{2}} = 3 - \frac{\pi}{2}.$

由定义知道，概率密度 $f(x)$ 具有以下性质：

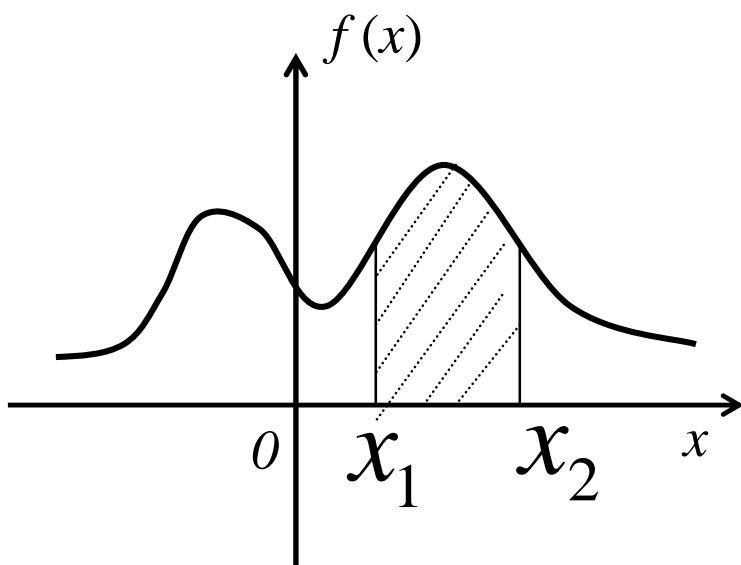
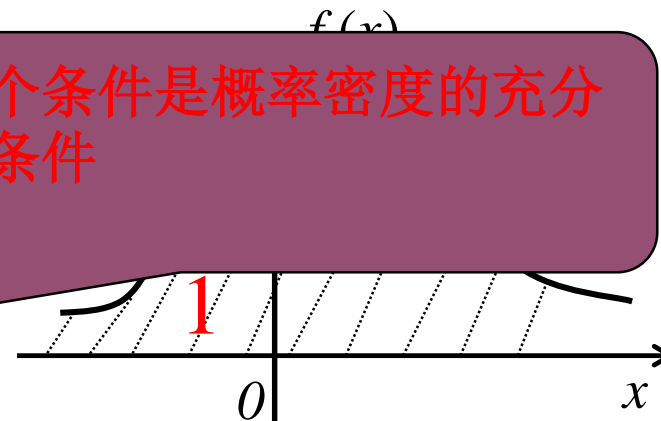
$$1^0 \quad f(x) \geq 0.$$

$$2^0 \quad \int_{-\infty}^{\infty} f(x)dx = 1.$$

$$3^0 \quad P\{x_1 < X \leq x_2\} = F(x_2) - F(x_1)$$

$$= \int_{x_1}^{x_2} f(x)dx. (x_1 \leq x_2)$$

前两个条件是概率密度的充分必要条件



4⁰ 若 $f(x)$ 在点 x 处连续，则有

$$F'(x) = f(x).$$

$$F(x) = \int_{-\infty}^x f(t)dt,$$

即

$$\begin{aligned} f(x) &= \lim_{\Delta x \rightarrow 0^+} \frac{F(x + \Delta x) - F(x)}{\Delta x} \\ &= \lim_{\Delta x \rightarrow 0^+} \frac{P\{x < X \leq x + \Delta x\}}{\Delta x} \end{aligned}$$

若不计高阶无穷小，有

$$P\{x < X \leq x + \Delta x\} \approx f(x)\Delta x.$$

注意

连续型随机变量密度函数的性质与离散型随机变量分布律的性质非常相似，但是，**密度函数不是概率！**

我们不能认为： $P\{X = a\} = f(a)$ ！

连续型随机变量的一个重要特点：

设 X 是连续型随机变量，对任意的实数 a ，

有
$$P\{X = a\} = 0$$

说明

(1) 由上述性质可知，对于连续型随机变量，我们关心它在某一点取值的问题没有太大的意义；我们所关心的是它在某一区间上取值的问题。

若已知连续型随机变量 X 的密度函数为 $f(x)$ ，则 X 在任意区间 G （ G 可以是开区间也可以是闭区间，或半开半闭区间；可以是有限区间也可以是无穷区间）上取值的概率为，

$$P\{X \in G\} = \int_G f(x) dx$$

此公式非常重要！

例 1

设 X 是连续型随机变量，其密度函数为

$$f(x) = \begin{cases} c(4x - 2x^2) & 0 < x < 2 \\ 0 & \text{其它} \end{cases}$$

求：(1) 常数 c ； (2) $P\{X > 1\}$.

解：(1) 由密度函数的性质

$$\int_{-\infty}^{+\infty} f(x) dx = 1$$

$$\begin{aligned} \text{得 } 1 &= \int_{-\infty}^{+\infty} f(x)dx = \int_0^2 c(4x - 2x^2)dx = c \left(2x^2 - \frac{2}{3}x^3 \right) \Big|_0^2 \\ &= \frac{8}{3}c \end{aligned}$$

$$\text{所以, } c = \frac{3}{8} \quad P\{X \in G\} = \int_G f(x)dx$$

$$\begin{aligned} (2) \quad P\{X > 1\} &= \int_1^{+\infty} f(x)dx = \int_1^2 \frac{3}{8}(4x - 2x^2)dx \\ &= \frac{3}{8} \left(2x^2 - \frac{2}{3}x^3 \right) \Big|_1^2 = \frac{1}{2} \end{aligned}$$

二、一些常用的连续型随机变量

1) 均匀分布

若随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其它} \end{cases}$$

则称随机变量 X 服从区间 $[a, b]$ 上的均匀分布。

记作 $X \sim U[a, b]$

密度函数的验证

设 $X \sim U[a, b]$, $f(x)$ 是其密度函数, 则有

(1) 对任意的 x , 有 $f(x) \geq 0$;

$$\begin{aligned} (2) \quad \int_{-\infty}^{+\infty} f(x) dx &= \int_{-\infty}^a f(x) dx + \int_a^b f(x) dx + \int_b^{+\infty} f(x) dx \\ &= \int_a^b \frac{1}{b-a} dx = 1. \end{aligned}$$

由此可知, $f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{其它} \end{cases}$ 确是密度函数.

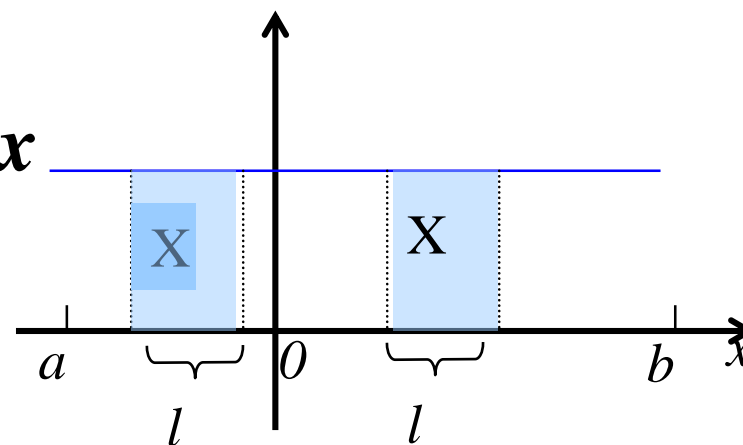
均匀分布的概率背景

如果随机变量 X 服从区间 $[a, b]$ 上的均匀分布，则随机变量 X 在区间 $[a, b]$ 上的任意一个子区间上取值的概率与该子区间的长度成正比，而与子区间的位置无关

这时，可以认为随机变量 X 在区间 $[a, b]$ 上取值是等可能的

$$P\{c < X \leq c + l\} = \int_c^{c+l} f(x)dx$$

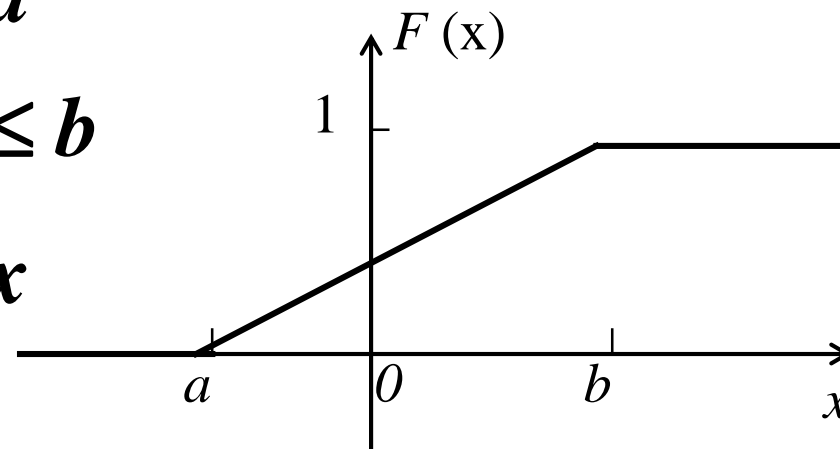
$$= \int_c^{c+l} \frac{1}{b-a} dx = \frac{l}{b-a}.$$



均匀分布的分布函数

若随机变量 X 服从区间 $[a, b]$ 上的均匀分布，
则 X 的分布函数为

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & b < x \end{cases}$$



例2 设公共汽车站从上午7时起每隔15分钟来一班车,如果某乘客到达此站的时间是 7:00 到7:30之间的均匀随机变量. 试求该乘客候车时间不超过5分钟的概率.

解: 设该乘客于7时 X 分到达此站.

则 X 服从区间 $[0, 30]$ 上的均匀分布.

$$\text{其密度函数为 } f(x) = \begin{cases} \frac{1}{30} & 0 \leq x \leq 30 \\ 0 & \text{其它} \end{cases}$$

令： $B = \{ \text{候车时间不超过5分钟} \}$

则 $P(B) = P\{10 \leq X \leq 15\} + P\{25 \leq X \leq 30\}$

$$= \int_{10}^{15} \frac{1}{30} dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{3}$$

2) 指数分布

如果随机变量 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

其中 $\lambda > 0$ 为常数，则称随机变量服从参数为 λ 的指数分布。

指数分布的分布函数

若随机变量 X 服从参数 λ 指数分布，则 X 的分布函数为

$$F(x) = \begin{cases} 0 & x \leq 0 \\ 1 - e^{-\lambda x} & x > 0 \end{cases}$$

例 3

设打一次电话所用的时间 X （单位：分钟）是以 $\lambda = \frac{1}{10}$ 为参数的指数随机变量。如果某人刚好在你前面走进公用电话间，求你需等待 0 分钟到 20 分钟之间的概率。

解：

$$X \text{ 的密度函数为 } f(x) = \begin{cases} \frac{1}{10} e^{-\frac{x}{10}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

令： $B = \{ \text{等待时间为} 10 \sim 20 \text{分钟} \}$

则 $P(B) = P\{10 \leq X \leq 20\}$

$$\begin{aligned}
 &= \int_{10}^{20} \frac{1}{10} e^{-\frac{x}{10}} dx = -e^{-\frac{x}{10}} \Big|_{10}^{20} \\
 &= e^{-1} - e^{-2} = 0.2325
 \end{aligned}$$

3) 正 态 分 布

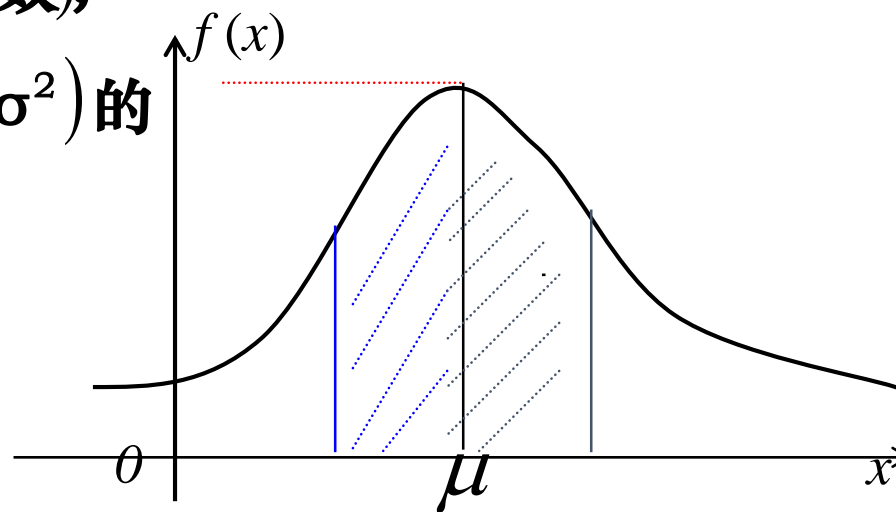
如果连续型随机变量 X 的密度函数为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

(其中 $-\infty < \mu < +\infty$, $\sigma > 0$ 为参数),

则称随机变量 X 服从参数为 (μ, σ^2) 的正态分布. 记作

$$X \sim N(\mu, \sigma^2)$$



标准正态分布

若 $\mu = 0$, $\sigma = 1$, 我们称 $N(0, 1)$ 为标准正态分布

标准正态分布的密度函数为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty < x < +\infty)$$

正态分布密度函数的图形性质

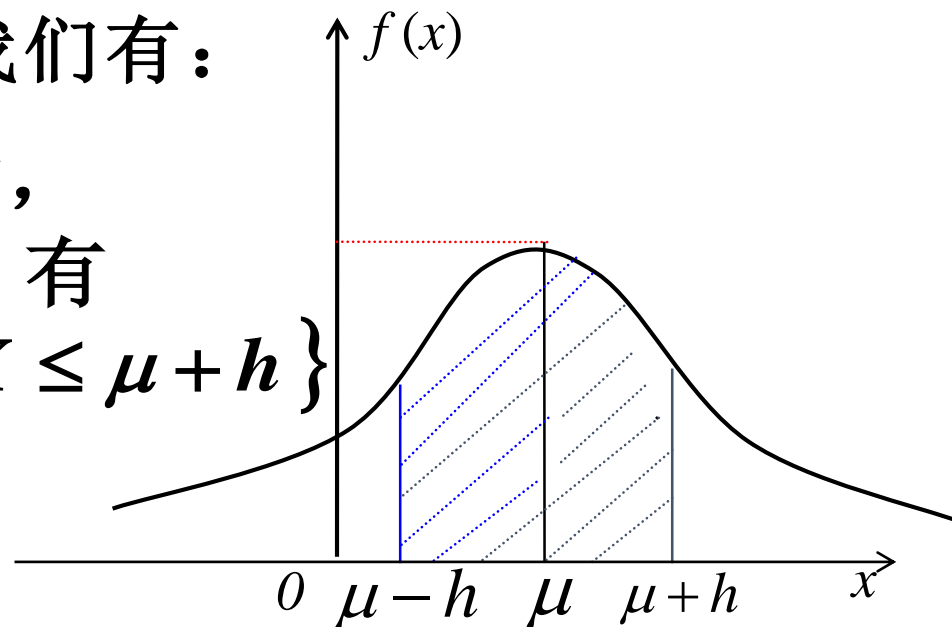
对于正态分布的密度函数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (-\infty < x < +\infty)$$

由高等数学中的知识,我们有:

(1) 曲线关于直线 $x = \mu$ 对称,
这表明: 对于任意的 $h > 0$, 有

$$P\{\mu - h < X \leq \mu\} = P\{\mu < X \leq \mu + h\}$$

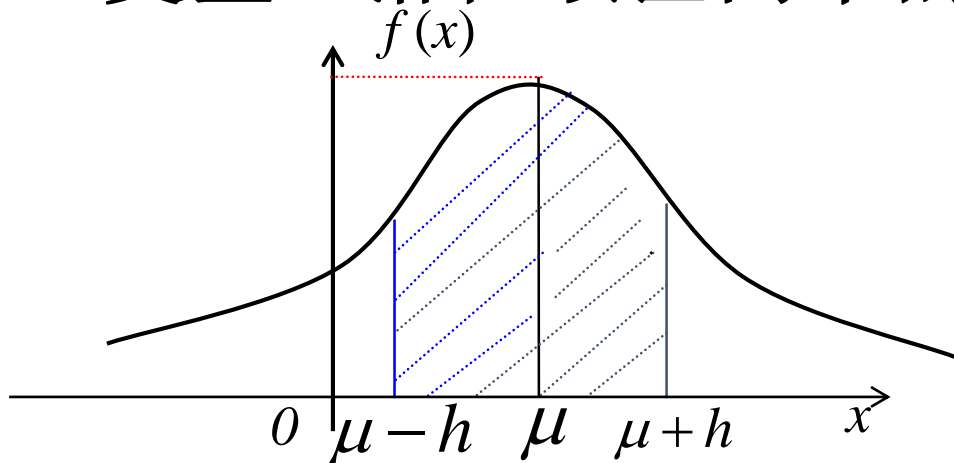


正态分布密度函数的图形性质（续）

(2) 当 $x = \mu$ 时， $f(x)$ 取到最大值

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$$

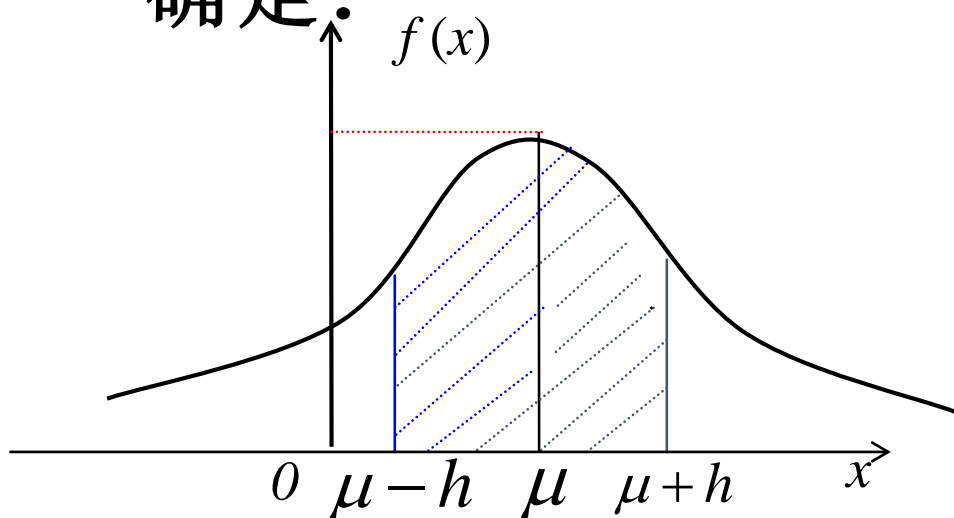
x 离 μ 越远， $f(x)$ 的值就越小。这表明，对于同样长度的区间，当区间离 μ 越远时，随机变量 X 落在该区间中的概率就越小。



正态分布密度函数的图形性质（续）

(3) 曲线 $y = f(x)$ 在 $x = \mu \pm \sigma$ 处有拐点；
曲线 $y = f(x)$ 以 Ox 轴为渐近线。

(4) 若 σ 固定，而改变 μ 的值，则 $f(x)$ 的图形沿 x 轴平行移动，但不改变其形状。因此 $y = f(x)$ 图形的位置完全由参数 μ 所确定。

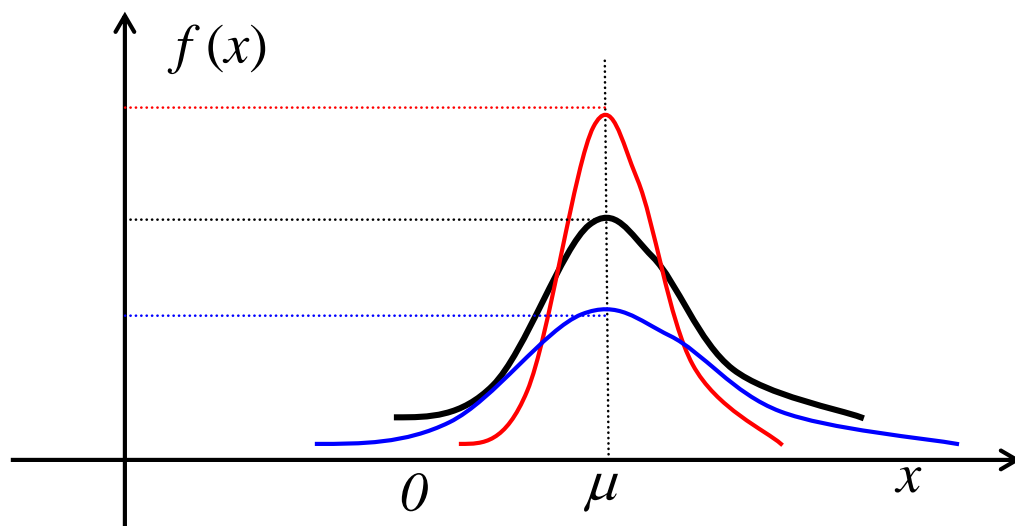


正态分布密度函数的图形性质（续）

(5) 若 μ 固定，而改变 σ 的值，由于 $f(x)$ 的最大值为

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma}$$

可知，当 σ 越小时， $y = f(x)$ 图形越陡，因而 X 落在 μ 附近的概率越大；反之 当 σ 越大时， $y = f(x)$ 的图形越平坦，这表 X 的取值越分散。



正态分布的重要性

正态分布是概率论中最重要的分布，这可以由以下情形加以说明：

- (1) 正态分布是自然界及工程技术中最常见的分布之一，大量的随机现象都是服从或近似服从正态分布的。可以证明，如果一个随机指标受到诸多因素的影响，但其中任何一个因素都不起决定性作用，则该随机指标一定服从或近似服从正态分布。
- (2) 正态分布有许多良好的性质，这些性质是其它许多分布所不具备的。
- (3) 正态分布可以作为许多分布的近似分布。

标准正态分布的计算

如果随机变量 $X \sim N(0, 1)$ ，则其密度函数为

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (-\infty, +\infty)$$

其分布函数为

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad (-\infty < x < +\infty)$$

教科书上都附有标准正态分布表,由此可得 $\Phi(x)$ 值.

标准正态分布的计算（续）

对于 $x \geq 0$ 我们可直接查表求出 $\Phi(x) = P\{X \leq x\}$

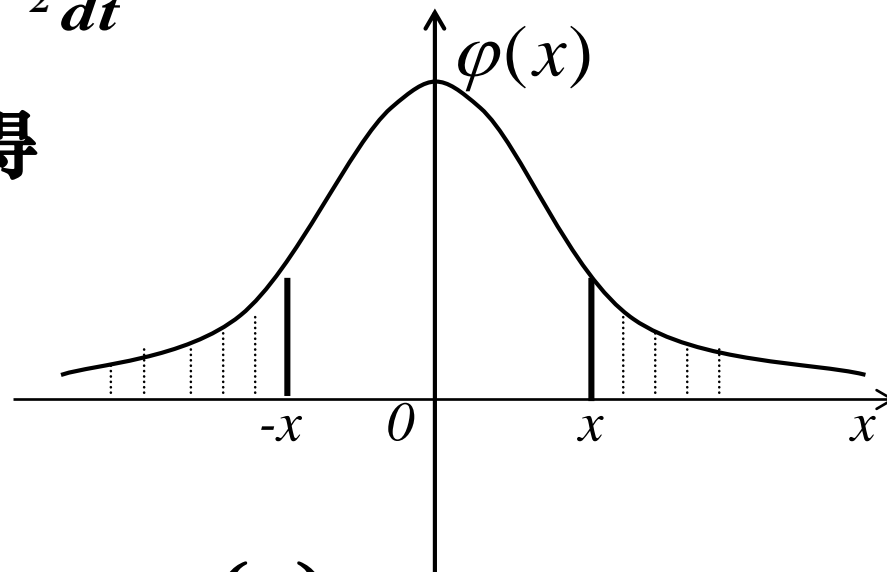
如果 $x < 0$ ，我们可由公式

$$\Phi(-x) = \int_{-\infty}^{-x} \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{-x} e^{-\frac{t^2}{2}} dt$$

作变换 $t = -u$ ， $dt = -du$ ，得

$$\Phi(-x) = -\frac{1}{\sqrt{2\pi}} \int_{+\infty}^x e^{-\frac{u^2}{2}} du$$

$$= \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{-\frac{u^2}{2}} du = 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du = 1 - \Phi(x)$$



一般正态分布的计算

设 $X \sim N(\mu, \sigma^2)$, 则 $Y = \frac{X - \mu}{\sigma} \sim N(0, 1)$

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\left\{\frac{X - \mu}{\sigma} \leq y\right\} \\ &= P\{X \leq \mu + \sigma y\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{\mu + \sigma y} e^{-\frac{(t - \mu)^2}{2\sigma^2}} dt \end{aligned}$$

作变换 $u = \frac{t - \mu}{\sigma}$, 则 $du = \frac{dt}{\sigma}$, 代入上式, 得

$$F_Y(y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^y e^{-\frac{u^2}{2}} du = \Phi(y)$$

$$\therefore F_X(x) = P\{X \leq x\} = P\left\{\frac{X - \mu}{\sigma} \leq \frac{x - \mu}{\sigma}\right\} = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

一般正态分布的计算（续）

其中， $\Phi(x)$ 是标准正态分

该公式给出了一般正态分布分布函数值的求法

$$\therefore F_x(x) = \Phi\left(\frac{x - \mu}{\sigma}\right)$$

故对任意的 $a < b$, 有

$$P\{a < X < b\} = \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$

例 4 设随机变量 $X \sim N(0, 1)$, 试求:
(1) $P\{1 \leq X < 2\}$; (2) $P\{-1 < X < 2\}$.

解:

$$\begin{aligned} (1) \quad P\{1 \leq X < 2\} &= \Phi(2) - \Phi(1) \\ &= 0.97725 - 0.84134 = 0.13591 \end{aligned}$$

$$\begin{aligned} (2) \quad P\{-1 \leq X < 2\} &= \Phi(2) - \Phi(-1) \\ &= \Phi(2) - [1 - \Phi(1)] \\ &= 0.97725 - 1 + 0.84134 = 0.81859 \end{aligned}$$

设 $X \sim N(0, 1)$, 若 z_α 满足条件

$$P\{X > z_\alpha\} = \alpha, 0 < \alpha < 1,$$

则称点 z_α 为标准正态分布的上 α 分位点。

查表可知

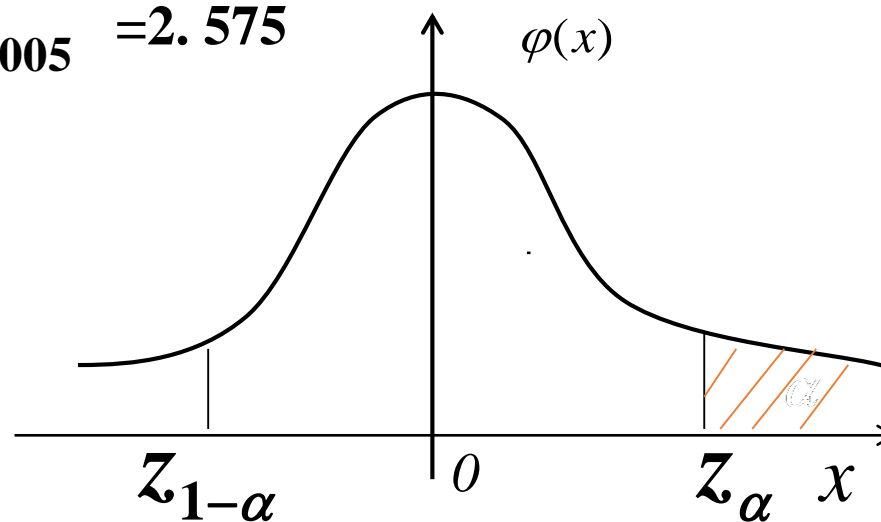
$$z_{0.05} = 1.645$$

$$z_{0.005} = 2.575$$

注: $z_{1-\alpha} = -z_\alpha,$

$$z_{0.95} = -1.645$$

$$z_{0.995} = -2.575$$



4) Γ -分布.

如果连续型随机变量 X 的密度函数为

$$f(x) = \begin{cases} \frac{\lambda^r}{\Gamma(r)} x^{r-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

(其中 $r > 0, \lambda > 0$ 为参数)

则称随机变量 X 服从参数为 (r, λ) 的 Γ -分布.

记作: $X \sim \Gamma(r, \lambda)$

Γ-函数

Γ-函数的定义：

$$\Gamma(r) = \int_0^{+\infty} x^{r-1} e^{-x} dx$$

Γ-函数的定义域： $(0, +\infty)$.

Γ-函数的性质： $\Gamma(r+1) = r\Gamma(r)$.

$$\Gamma(1) = 1, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}.$$

如果 n 为自然数，则 $\Gamma(n) = (n-1)!$.

说明：

如果 $r=1$ ，则由 $\Gamma(1)=1$ ，得 $f(x)=\begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$

这正是参数为 λ 的指数分布。

这说明指数分布是 Γ -分布的一个特例。

如果 $r=n$ ，由 $\Gamma(n)=(n-1)!$ 得

$$f(x)=\begin{cases} \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

我们称此分布为 *Erlang* 分布，它是排队论中重要的分布之一。

说明：

如果 $r = \frac{n}{2}$, $\lambda = \frac{1}{2}$, 其中 n 为自然数, 则有

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} x^{\frac{n}{2}-1} e^{-\frac{x}{2}} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

我们称此分布为自由度为 n 的 χ^2 -分布, 记作 $\chi^2(n)$.
它是数理统计学中重要的分布之一.

期 望

期望(mean)：也就是均值，是概率加权下的“平均值”，是每次可能结果的概率乘以其结果的总和，反映的实随机变量平均取值大小。

常用符号 μ 表示

连续性 $E(X) = \int_{-\infty}^{\infty} xf(x)dx$

离散型 $E(X) = \sum_i x_i p_i$

期望

X	2	4	6	8	10
P(x)	0.2	0.2	0.2	0.2	0.2

$$\begin{aligned}
 E(X) &= \sum_i x_i p_i \\
 &= 2 * 0.2 + 4 * 0.2 + 6 * 0.2 + 8 * 0.2 + 10 * 0.2 \\
 &= 6
 \end{aligned}$$

期望

假设C为一个常数，X和Y为两个随机变量，那么期望有以下性质：

$$E(C) = C \quad E(CX) = CE(X)$$

$$E(X + Y) = E(X) + E(Y)$$

如果X和Y相互独立，那么 $E(XY) = E(X)E(Y)$

如果 $E(XY) = E(X)E(Y)$ ，那么X和Y不相关

甲乙两个人赌博，假设两人获胜的概率相等，比赛规则是先胜三局者为赢家，可以获得100元的奖励。当比赛进行了三局的时候，其中甲胜了两局，乙胜了一局，这个时候由于某些原因中止了比赛，请问如何分配这100元才比较公平？

$$P(\text{甲}) = P(\text{乙}) = \frac{1}{2}$$

$$P(\text{甲} = \text{赢}) = P(\text{赢} | 4) + P(\text{赢} | 5, \text{输} | 4) = \frac{1}{2} + \frac{1}{2} * \left(1 - \frac{1}{2}\right) = \frac{3}{4}$$

$$P(\text{乙} = \text{赢}) = P(\text{赢} | 5, \text{赢} | 4) = \frac{1}{2} * \frac{1}{2} = \frac{1}{4}$$

$$E(\text{甲}) = 100 * P(\text{甲} = \text{赢}) = 100 * \frac{3}{4} = 75$$

$$E(\text{乙}) = 100 * P(\text{乙} = \text{赢}) = 100 * \frac{1}{4} = 25$$

期望

某城市有10万个家庭，没有孩子的家庭有1000个，有一个孩子的有9万个，有两个孩子的家庭有6000个，有三个孩子的家庭有3000个，问此城市一个家庭平均有小孩多少个？

X	0	1	2	3
P	0.01	0.9	0.06	0.03

$$E(X) = \sum_{i=0}^3 x_i p(x_i) = 0 * 0.01 + 1 * 0.9 + 2 * 0.06 + 3 * 0.03 = 1.11$$

例3 甲、乙两人射击，他们的射击水平由下表给出：

X ：甲击中的环数； Y ：乙击中的环数；

X	8	9	10
P	0.1	0.3	0.6
Y	8	9	10
P	0.2	0.5	0.3

试问哪一个人的射击水平较高？

解：甲、乙的平均环数可写为

$$EX = 8 \times 0.1 + 9 \times 0.3 + 10 \times 0.6 = 9.5$$

$$EY = 8 \times 0.2 + 9 \times 0.5 + 10 \times 0.3 = 9.1$$

因此，从平均环数上看，甲的射击水平要比乙的好。

方差

方差(variance)是衡量随机变量或一组数据时离散程度的度量，是用来度量随机变量和其数学期望之间的偏离程度。即方差是衡量数据源数据和期望均值相差的度量值。

$$Var(X) = D(X) = \sigma^2 = \frac{\sum (X - \mu)^2}{N}$$

$$D(X) = \sum_{i=1}^n p_i \cdot (x_i - \mu)^2$$

$$D(X) = \int_a^b (x - \mu)^2 f(x) dx$$

$$D(X) = E((X - E(X))^2) = E(X^2) - (E(X))^2$$

方差

X	2	4	6	8	10
P(x)	0.2	0.2	0.2	0.2	0.2

$$E(X) = 6 \quad E(X^2) = 44$$

$$\begin{aligned}
 D(X) &= E(X^2) - (E(X))^2 \\
 &= 44 - 6^2 \\
 &= 8
 \end{aligned}$$

例1 甲、乙两人射击，他们的射击水平由下表给出：

X : 甲击中的环数；

Y : 乙击中的环数；

X	8	9	10
P	0.3	0.2	0.5
Y	8	9	10
P	0.2	0.4	0.4

试问哪一个人的射击水平较高？

解：比较两个人的平均环数.

甲的平均环数为

$$EX = 8 \times 0.3 + 9 \times 0.2 + 10 \times 0.5 = 9.2 \quad (\text{环})$$

乙的平均环数为

$$EY = 8 \times 0.2 + 9 \times 0.4 + 10 \times 0.4 = 9.2 \quad (\text{环})$$

因此，从平均环数上看，甲乙两人的射击水平是一样的，但两个人射击环数的方差分别为

$$\begin{aligned}DX &= (8-9.2)^2 \times 0.3 + (9-9.2)^2 \times 0.2 + (10-9.2)^2 \times 0.5 \\ &= 0.76\end{aligned}$$

$$\begin{aligned}DY &= (8-9.2)^2 \times 0.2 + (9-9.2)^2 \times 0.4 + (10-9.2)^2 \times 0.4 \\ &= 0.624\end{aligned}$$

由于 $DY < DX$,

这表明乙的射击水平比甲稳定.

方差

假设C为一个常数，X和Y实两个随机变量，那么方差有以下性质：

$$D(C) = 0 \quad D(CX) = C^2 D(X) \quad D(C + X) = D(X)$$

$$D(X \pm Y) = D(X) + D(Y) \pm 2Cov(X, Y)$$

$$\text{协方差 } Cov(X, Y) = E\{(X - E(X)) \cdot (Y - E(Y))\}$$

如果X和Y不相关，那么 $D(X \pm Y) = D(X) + D(Y)$

常见分布

分 布	参 数	数学期望	方差
两点分布	$0 < p < 1$	p	$p(1-p)$
二项分布	$n \geq 1,$ $0 < p < 1$	np	$np(1-p)$
泊松分布	$\lambda > 0$	λ	λ
均匀分布	$a < b$	$(a+b)/2$	$(b-a)^2/12$
指数分布	$\theta > 0$	θ	θ^2
正态分布	$\mu, \sigma > 0$	μ	σ^2

标准差

标准差(Standard Deviation)是离均值平方的算术平均数的平方根，用符号 σ 表示，其实标准差就是方差的算术平方根。

标准差和方差都是测量离散趋势的最重要、最常见的指标。标准差和方差的不同点在于，标准差和变量的计算单位是相同的，比方差清楚，因此在很多分析的时候使用的是标准差。

$$\sigma = \sqrt{D(X)} = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$

标准差

X1	2	4	6	8	10
P(x1)	0.2	0.2	0.2	0.2	0.2

X2	4	5	6	7	8
P(x2)	0.2	0.2	0.2	0.2	0.2

$$D(X_1) = 8$$

$$\sigma_1 = \sqrt{D(X_1)} = \sqrt{8} = 2.8284$$

$$D(X_2) = 2$$

$$\sigma_2 = \sqrt{D(X_2)} = \sqrt{2} = 1.4142$$

案例1

有甲乙两个单位愿意聘用你，而你能够获得的信息如下，请根据工资待遇的差异情况，您选择哪家单位？为什么？

甲单位不同职位月工资 x_1 元	1200	1400	1600	1800
获取该职位的概率 P_1	0.4	0.3	0.2	0.1

乙单位不同职位月工资 x_2 元	1000	1400	1800	2200
获取该职位的概率 P_2	0.4	0.3	0.2	0.1

$$E(X_1) = 1400 \qquad E(X_2) = 1400$$

$$D(X_1) = 40000 \qquad D(X_2) = 160000$$

案例2

已知随机变量X的分布列如下，分别求 $E(X)$ 、 $E(2X+5)$ 、 $D(X)$ 、 $\sigma(X)$ 的值

X	-2	1	3
P	0.16	0.44	0.40

$$E(X) = -2 * 0.16 + 1 * 0.44 + 3 * 0.40 = 1.32$$

$$E(2X + 5) = 2E(X) + 5 = 2 * 1.32 + 5$$

$$D(X) = E(X^2) - (E(X))^2 = (-2)^2 * 0.16 + 1^2 * 0.44 + 3^2 * 0.40 - 1.32^2 = 2.9376$$

$$\sigma(X) = \sqrt{D(X)} = \sqrt{2.9376} \approx 1.7139$$

协方差

协方差常用于衡量两个变量的总体误差；当两个变量相同的情况下，协方差其实就是方差。

如果X和Y是统计独立的，那么二者之间的协方差为零。但是如果协方差为零，那么X和Y是不相关的。

$$\begin{aligned} Cov(X, Y) &= E[(X - E(X)) \cdot (Y - E(Y))] \\ &= E[XY - XE(Y) - YE(X) + E(X)E(Y)] \\ &= E(XY) - E(X)E(Y) \end{aligned}$$

协方差

假设C为一个常数，X和Y实两个随机变量，那么协方差有性质如下所示：

$$Cov(X, Y) = Cov(Y, X)$$

$$Cov(aX, bY) = abCov(X, Y)$$

$$Cov(X_1 + X_2, Y) = Cov(X_1, Y) + Cov(X_2, Y)$$

协方差

协方差是两个随机变量具有相同方向变化趋势的度量

1. 若 $\text{Cov}(X, Y) > 0$, 则X和Y的变化趋势相同
2. 若 $\text{Cov}(X, Y) < 0$, 则X和Y的变化趋势相反
3. 若 $\text{Cov}(X, Y) = 0$, 则X和Y不相关, 也就是变化没有什么相关性

协方差矩阵

对于n个随机向量($X_1, X_2, X_3 \dots X_n$), 任意两个元素 X_i 和 X_j 都可以得到一个协方差, 从而形成一个 $n \times n$ 的矩阵, 该矩阵就叫做协方差矩阵, 协方差矩阵为对称矩阵。

$$c_{ij} = E\{[X_i - E(X_i)][X_j - E(X_j)]\} = Cov(X_i, X_j)$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{nn} \end{bmatrix}$$

协方差的上界

$$\text{Cov}(X, Y) = E[(X - E(X)) \cdot (Y - E(Y))]$$

$$\begin{aligned} \text{Cov}^2(X, Y) &= E^2[(X - E(X)) \cdot (Y - E(Y))] \\ &\leq E\{[(X - E(X)) \cdot (Y - E(Y))]^2\} \\ &= E[(X - E(X))^2 (Y - E(Y))^2] \\ &\leq E[(X - E(X))^2] E[(Y - E(Y))^2] \\ &= \text{Var}(X) \text{Var}(Y) \end{aligned}$$

Pearson相关系数

协方差可以描述X和Y的相关程度，但是协方差的值和X/Y的值采用的是不同的量纲，导致协方差在数值上表现出比较大的差异，因此可以引入相关系数来表示X和Y的相关性。

$$\rho(X,Y)=\frac{Cov(X,Y)}{\sqrt{D(X)}\cdot\sqrt{D(Y)}}$$

当 $\rho(X,Y)=0$ 的时候，称X和Y不线性相关。

Pearson相关系数取值范围为[-1,1]

绝对值范围	含义
0.8-1.0	极强相关
0.6-0.8	强相关
0.4-0.6	中等程度相关
0.2-0.4	弱相关
0-0.2	极弱相关或无相关

中心矩、原点矩

假设 X 和 Y 是随机变量，若 $E(X^k), k=1, 2, \dots$ 存在，则称它为 X 的 k 阶原点矩，简称 k 阶矩。

若 $E\{[X-E(X)]^k\}, k=1, 2, \dots$ 存在，则称它为 X 的 k 阶中心矩。

若 $E\{[X-c]^k\}, k=1, 2, \dots$ 存在，则称它为 X 关于点 c 的 k 阶矩。

若 $E\{X^k Y^p\}, k, p=1, 2, \dots$ 存在，则称它为 X 和 Y 的 $k+p$ 阶混合原点矩。

若 $E\{[X-E(X)]^k [Y-E(Y)]^p\}, k, p=1, 2, \dots$ 存在，则称它为 X 和 Y 的 $k+p$ 阶混合中心矩。

中心矩、原点矩

X 的数学期望 $E(X)$ 是 X 的一阶原点矩。

X 的方差 $D(X)$ 是 X 的二阶中心矩。

X 和 Y 的协方差 $\text{Cov}(X, Y)$ 是 X 和 Y 的二阶混合中心矩。

峰度

峰度(peakedness; kurtosis)又称峰态系数。表示了概率密度分布曲线在平均值处峰值高低的特征数，直观来讲，峰度反映的是峰部的尖度。

样本的峰度是和正态分布相比较而言的统计量，如果峰度值大于三，那么峰的形状比较尖，比正态分布峰要陡峭。反之亦然。

峰度计算公式：随机变量的四阶中心矩与方差平方的比值。

$$kurtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^4}{(N-1) \cdot s^4}$$

偏度

偏度系数(skewness)是描述分布偏离对称性程度的一个特征数，当分布左右对称的时候，偏度系数为0，当偏度系数大于0时候，即重尾在右侧时，该分布为右偏；当偏度系数小于0时候，即重尾在左侧时，该分布为左偏。

偏度计算公式：随机变量的三阶中心矩与样本的平均离均差立方和的比值。

$$kurtosis = \frac{\sum_{i=1}^N (x_i - \bar{x})^3}{(N-1) \bullet s^3}$$

切比雪夫不等式/切比雪夫定理

设随机变量 X 的期望为 μ ，方差为 σ^2 ，对于任意的正数 ε ，有：

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

切比雪夫不等式的含义是：DX(方差)越小，时间 $\{|X - \mu| < \varepsilon\}$ 发生的概率就越大，即： X 取的值基本上集中在期望 μ 附近

大数定律

设随机变量 X_1, X_2, \dots, X_n 是一列相互独立的随机变量(或者两两不相关), 并且分别存在期望 $E(X_k)$ 和方差 $D(X_k)$, 对于任意小的正数 ε , 有:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| < \varepsilon \right\} = 1$$

当具有相同期望 μ 和方差为 σ^2 的时候, 对随机变量的均值: $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$

则有: $\lim_{n \rightarrow \infty} P \{ |Y_n - \mu| < \varepsilon \} = 1$

大数定律

大数定律的意义：随着样本容量 n 的增加，样本平均数将接近于总体平均数(期望 μ)，所以在统计推断中，一般都会使用样本平均数估计总体平均数的值。

也就是我们会使用一部分样本的平均值来代替整体样本的期望/均值，出现偏差的可能是存在的，但是当 n 足够大的时候，偏差的可能性是非常小的，当 n 无限大的时候，这种可能性的概率基本为0。

大数定律的主要作用就是为使用频率来估计概率提供了理论支持。

中心极限定理

中心极限定理的意义：设从均值为 μ 、方差为 σ^2 有限的任意一个总体中抽取样本量为 n 的样本，当 n 充分大时，样本均值的抽样分布近似服从均值为 μ/n 、方差为 σ^2/n 的正态分布。

中心极限定理就是一般在同分布的情况下，抽样样本值的和在总体数量趋于无穷时的极限分布近似与正态分布。

中心极限定理

中心极限定理(Central Limit Theorem)；假设 $\{X_n\}$ 为独立同分布的随机变量序列，并具有相同的期望 μ 和方差为 σ^2 ，则 $\{X_n\}$ 服从中心极限定理，且 Z_n 为随机序列 $\{X_n\}$ 的规范和：

$$Y_n = X_1 + X_2 + \dots + X_n = \sum_{i=1}^n X_i$$

$$Z_n = \frac{Y_n - E(Y_n)}{\sqrt{D(Y_n)}} = \frac{Y_n - n\mu}{\sqrt{n\sigma^2}} \rightarrow N(n\mu, n\sigma^2)$$

参数估计

一、点估计

设总体 X 的分布函数 $F(x; \theta)$ 的形式为已知, θ 是待估参数。 X_1, \dots, X_n 是 X 的一个样本, x_1, \dots, x_n 是相应的样本值。

构造一个适当的统计量 $\hat{\theta}(X_1, \dots, X_n)$, 用它的观察值 $\hat{\theta}(x_1, \dots, x_n)$ 来估计未知参数 θ 。

我们称 $\hat{\theta}(X_1, \dots, X_n)$ 为 θ 的估计量;

称 $\hat{\theta}(x_1, \dots, x_n)$ 为 θ 的估计值。

这种对未知参数进行数值估计的问题就是点估计问题。

注意:

(1) 估计量与估计值有着本质的不同:

估计量是统计量, 因而它是随机变量(一维或多维); 而估计值则是一维或多维数组.

(2) 在不引起混淆的情况下, 我们统称估计量与估计值为未知参数 θ 的估计.

二、矩估计

设 X 为连续型随机变量，其概率密度为

$$f(x; \theta_1, \dots, \theta_k),$$

X 为离散型随机变量，其分布列为

$$P\{X = x\} = p(x; \theta_1, \dots, \theta_k),$$

其中 $\theta_1, \dots, \theta_k$ 是待估参数， X_1, \dots, X_n 为来自 X 的样本。

设 $EX^l = \mu_l$ 存在, $l = 1, 2, \dots, k$

则 $\mu_l = \mu_l(\theta_1, \dots, \theta_k), l = 1, 2, \dots, k.$

令 $A_l = \mu_l, l = 1, \dots, k$, 其中 $A_l = \frac{1}{n} \sum_{i=1}^n X_i^l$

这是包含 k 个未知参数 $\theta_1, \dots, \theta_k$ 的联立方程组:

$$\begin{cases} A_1 = \mu_1(\theta_1, \theta_2, \dots, \theta_k) \\ A_2 = \mu_2(\theta_1, \theta_2, \dots, \theta_k) \\ \dots \dots \dots \dots \\ A_k = \mu_k(\theta_1, \theta_2, \dots, \theta_k) \end{cases}$$

从中解出方程组的解记为 $\hat{\theta}_1, \dots, \hat{\theta}_k$,即

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_n) \\ \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_n) \\ \dots \dots \dots \dots \\ \hat{\theta}_k = \hat{\theta}_k(X_1, X_2, \dots, X_n) \end{cases}$$

用 $\hat{\theta}_1, \dots, \hat{\theta}_k$ 分别作为 $\theta_1, \dots, \theta_k$ 的估计量,

这种求估计量的方法称为矩估计法

这种估计量称为**矩估计量**；矩估计量的观察值称为**矩估计值**。

矩法原理：由辛钦大数定律知

$$A_l = \frac{1}{n} \sum_{i=1}^n X_i^l \xrightarrow{P} \mu_l, \quad l = 1, 2, \dots, k.$$

所以我们令 $A_l = \mu_l, \quad l = 1, \dots, k$, 用 A_l 估计 μ_l .

矩法求估计量的步骤:

1) 求 $\mu_1 = EX$ ($\mu_2 = EX^2$);

2) 令 $A_1 = \mu_1$ ($A_2 = \mu_2$);

3) 解上面方程 (组), 得

$$\begin{aligned} \hat{\theta}_1 &= \hat{\theta}_1(X_1, \dots, X_n) \\ (\hat{\theta}_2 &= \hat{\theta}_2(X_1, \dots, X_n)). \end{aligned}$$

例 1 设某炸药厂一天中发生着火现象的次数 X 服从参数为 λ 的泊松分布， λ 未知，有以下样本值试估计参数 λ （用矩法）。

着火的次数 k	0	1	2	3	4	5	6	
发生 k 次着火天数 n_k	75	90	54	22	6	2	1	$\sum = 250$

解: $\mu_1 = EX = \lambda, \quad A_1 = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X}$

令 $\bar{X} = \lambda,$

则 $\hat{\lambda} = \bar{x} = \frac{1}{250} (0 \times 75 + 1 \times 90 + \cdots + 6 \times 1) = 1.22$

所以估计值 $\hat{\lambda} = 1.22$ 。

例2

设总体 $X \sim U[a, b]$, a, b 未知, X_1, \dots, X_n 是一个样本,
求: a, b 的矩估计量。

解:

$$\mu_1 = EX = \frac{a+b}{2},$$

$$\mu_2 = EX^2 = DX + (EX)^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

$$\text{令 } \frac{a+b}{2} = A_1$$

$$\text{即 } a+b = 2A_1,$$

$$\frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = A_2$$

$$b-a = \sqrt{12(A_2 - A_1^2)}$$

即 $a + b = 2A_1, \quad b - a = \sqrt{12(A_2 - A_1^2)}$

解得：

$$\hat{a} = A_1 - \sqrt{3(A_2 - A_1^2)} = \bar{X} - \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$\hat{b} = A_1 + \sqrt{3(A_2 - A_1^2)} = \bar{X} + \sqrt{\frac{3}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

$$A_2 - A_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \left(\sum_{i=1}^n X_i^2 - n\bar{X}^2 \right) = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

例3 设总体 X 的均值 μ ，方差 σ^2 都存在，且 $\sigma^2 > 0$ ，但 μ ， σ^2 未知，又设 X_1, \dots, X_n 是一个样本；求： μ, σ^2 的矩估计量。

解： $\mu_1 = EX = \mu$,

$$\mu_2 = EX^2 = DX + (EX)^2 = \sigma^2 + \mu^2$$

令 $\mu_1 = A_1$, $\mu_2 = A_2$,

即 $\mu = A_1$, $\sigma^2 + \mu^2 = A_2$,

所以 $\hat{\mu} = A_1 = \bar{X}$,

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

特别，若 $X \sim N(\mu, \sigma^2)$, μ, σ^2 未知；

$$\text{则 } \hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

例4 设总体 X 服从参数为 λ 的指数分布，其中 $\lambda > 0$ 未知， X_1, X_2, \dots, X_n 是从该总体中抽取的一个样本，试求参数 λ 的矩估计。

解： 总体 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \leq 0. \end{cases}$$

所以, $EX = \int_{-\infty}^{+\infty} xf(x)dx = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$

令

$$\bar{X} = \frac{1}{\lambda},$$

得参数 λ 的矩估计量为

$$\hat{\lambda} = \frac{1}{\bar{X}}.$$

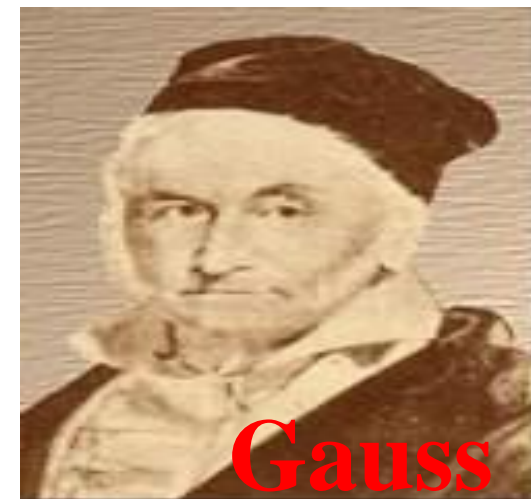
二、极大似然估计法

极大似然估计法是在**总体的分布类型已知**的条件下所使用的一种参数估计方法。

它首先是由德国数学家**高斯**在1821年提出的。

然而，这个方法常归功于英国统计学家**费歇**。

费歇在1922年重新发现了这一方法，并首先研究了这种方法的一些性质。



极大似然原理： 一个随机试验有若干个可能结

果A,B,C,...。若在一次试验中，结果A发生，

则一般认为试验条件对A最有利，即A发生的

概率 $P(A / \theta)$ 最大

如，甲 $\begin{cases} 99 & \text{红} \\ 1 & \text{黑} \end{cases}$ ，乙 $\begin{cases} 1 & \text{红} \\ 99 & \text{黑} \end{cases}$ ，任取1箱从中任取1球，

已知取到红球 问最有可能从何箱取

$$P(\text{红球}/\text{甲}) = 0.99 \quad P(\text{红球}/\text{乙}) = 0.01$$

自然，认为从甲箱取更合理

又如，兔龟赛跑，得第一名的最有可能是谁？

极大似然估计法：

(1) X ---离散型，已知 X 的分布

$$P(X = x) = p(x, \theta), \theta \text{未知}$$

样本 (X_1, X_2, \dots, X_n) 取到观测值 (x_1, x_2, \dots, x_n)

事件 A

$$P(A) = P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$$

独立

$$\underline{\underline{=}} P(X_1 = x_1)P(X_2 = x_2) \cdots P(X_n = x_n)$$

Xi与X 同分布

$$P(X = x_1)P(X = x_2) \cdots P(X = x_n)$$

$$= p(x_1, \theta)p(x_2, \theta) \cdots p(x_n, \theta)$$

$$= \prod_{i=1}^n p(x_i, \theta)$$

对给定的样本值 (x_1, x_2, \dots, x_n) , $\prod_{i=1}^n p(x_i, \theta)$

是参数 θ 的函数, 称为**似然函数**, 记做 $L(\theta)$.

$$\text{即 } L(\theta) = \prod_{i=1}^n p(x_i; \theta)$$

结构：n 项连乘，总体分布

$$p(x, \theta) \xrightarrow{\text{改}} p(x_i, \theta) \\ i = 1, 2, \dots, n$$

$P(A) = L(\theta)$, 随 θ 变而变, A 已经发生, 由极大似然原理, $L(\theta)$ 达到最大, 所以 θ 的最合理

估计值 $\hat{\theta}$ 应满足: $L(\hat{\theta})$ 为最大值

定义 对给定的样本值 x_1, x_2, \dots, x_n 若 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 满足

$$L(\hat{\theta}) = \max_{\theta} L(\theta)$$

称： $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 θ 的极大似然估计值
 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 θ 的极大似然估计量

如何求 $\hat{\theta}$? 即求 $L(\theta)$ 的最大值点问题

方法一： 若 $L(\theta)$ 为可导函数

$$\text{解方程 } \frac{dL(\theta)}{d\theta} = 0,$$

$$\text{得到 } \hat{\theta} = \hat{\theta}(X_1, X_2, \dots, X_n)$$

(1) $f(x) > 0$, $\ln[f(x)]$ 单调性相同, 从而 **最大值点相同**.

(2) $L(\theta) = \prod_{i=1}^n p(x_i; \theta)$ n 项连乘, 求导麻烦

$\ln[L(\theta)]$ n 项相加, 求导简单

对数似然函数

从而,

求的 $L(\theta)$ 最大值点就转为求 $\ln[L(\theta)]$ 的最大值点

方法二:

解方程 $\frac{d \ln[L(\theta)]}{d \theta} = 0$, 得到 $\hat{\theta}$

(2) 连续型总体似然函数的求法

设 X 为连续型总体，其概率密度为：

$$f(x; \theta) \quad \text{其中 } \theta \text{ 未知}$$

对来自总体的样本 (X_1, X_2, \dots, X_n) 其观测值为 (x_1, x_2, \dots, x_n) ，作为与总体 X 同分布且相互独立的 n 维随机变量，样本的联合概率密度为：

$$\begin{aligned} f(x_1, x_2, \dots, x_n) &= f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n) \\ &= f(x_1, \theta) \cdot f(x_2, \theta) \cdots f(x_n, \theta) = \prod_{i=1}^n f(x_i, \theta) \end{aligned}$$

于是，样本 (X_1, X_2, \dots, X_n) 落入点 (x_1, x_2, \dots, x_n)

邻域内的概率为 $\prod_{i=1}^n f(x_i, \theta) \Delta x_i$ 由极大似然原

理，最合理的 θ 的估计值 $\hat{\theta}$ 应该是使 $\prod_{i=1}^n f(x_i, \theta) \Delta x_i$ 达到最大，由于 Δx_i 是不依赖于 θ

的增量，所以我们只需求使

似然函数 $L(\theta) = \prod_{i=1}^n f(x_i, \theta)$ 达到最大

求 $\hat{\theta}$ 的步骤:

(1) 写出 $L(\theta)$

(2) 取对数 $\ln L(\theta)$

(3) 解方程 $\frac{d \ln[L(\theta)]}{d \theta} = 0$, 得到 $\hat{\theta}$

例1 一个罐子里装有黑球和白球，有放回地抽取 n 个球，发现有 k 个黑球。试求罐子里黑球数与白球数之比 R 的极大似然估计量。

解： 设罐中装有 a 只黑球 b 只白球，则 $R = \frac{a}{b}$ 。

设 $X_i = \begin{cases} 1, & \text{第} i \text{次摸到黑球;} \\ 0, & \text{第} i \text{次摸到白球} \end{cases} \quad i = 1, \dots, n.$

则 X_1, \dots, X_n 是总体 $X \sim b(1, p)$ 的样本，

其中 $p = P\{X_i = 1\} = \frac{a}{a+b} = \frac{R}{1+R}$ 。

$$\begin{aligned} \text{似然函数 } L(R) &= \prod_{i=1}^n p(x_i) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} \quad p = \frac{R}{1+R} \\ &= \prod_{i=1}^n \left(\frac{R}{1+R}\right)^{x_i} \left(1 - \frac{R}{1+R}\right)^{1-x_i} = \left(\frac{R}{1+R}\right)^{\sum_{i=1}^n x_i} \left(1 - \frac{R}{1+R}\right)^{n - \sum_{i=1}^n x_i}, \end{aligned}$$

$$\begin{aligned} \text{则 } \ln L &= \sum_{i=1}^n x_i \ln\left(\frac{R}{1+R}\right) + \left(n - \sum_{i=1}^n x_i\right) \ln\left(\frac{1}{1+R}\right) \\ &= \sum_{i=1}^n x_i [\ln R - \ln(1+R)] - \left(n - \sum_{i=1}^n x_i\right) \ln(1+R) \end{aligned}$$

$$\text{令 } \frac{d \ln L}{dR} = 0, \text{ 则 } \sum_{i=1}^n x_i \left(\frac{1}{R} - \frac{1}{1+R}\right) - \left(n - \sum_{i=1}^n x_i\right) \frac{1}{1+R} = 0$$

$$\text{解出 } \hat{R} = \frac{\sum_{i=1}^n x_i}{n - \sum_{i=1}^n x_i} = \frac{k}{n - k}.$$

例2 设 $X \sim N(\mu, \sigma^2)$; μ 已知, σ^2 为未知参数, x_1, \dots, x_n 是来自 X 的一个样本值求 σ^2 的极大似然估计量

解: X 的概率密度为:

$$f(x; \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

似然函数为:

$$\begin{aligned} L(\sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right\} \end{aligned}$$

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\frac{d \ln L}{d \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2$$

令： $\frac{d \ln L}{d \sigma^2} = 0$ ，

得似然方程 $-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$

解此方程，得 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$ ，

因此 σ^2 的极大似然估计量为 $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$ 。

例3 设 $X \sim N(\mu, \sigma^2)$; μ, σ^2 为未知参数, x_1, \dots, x_n 是来自 X 的一个样本值, 求 μ, σ^2 的极大似然估计量

解: X 的概率密度为:

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x - \mu)^2\right\}$$

似然函数为:

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\} \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2}} \\ \ln L &= -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 \end{aligned}$$

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

$$\text{令} \begin{cases} \frac{\partial \ln L}{\partial \mu} = 0 \\ \frac{\partial \ln L}{\partial \sigma^2} = 0 \end{cases} \quad \text{即:} \begin{cases} \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \\ -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0 \end{cases}$$

$$\text{解得: } \hat{\mu} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

故 μ, σ^2 的极大似然估计量为

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

参考文献

1. 盛骤，概率论与数理统计，浙江大学出版社，2008
2. 龙永红，概率论与数理统计，高等教育出版社，2013
3. 魏宗舒，概率论与数理统计教程，高等教育出版社，2013



THANK YOU

上海育创网络科技有限公司