《计算机模拟》

第2讲 – 概率基础

胡贤良
浙江大学数学科学学院

本讲内容

1. 概率论公理化

2. 随机变量初步

3. 案例：贝叶斯分类

把概率论建立在严格的逻辑基础上，

探索一直持续了3个世纪！

1. 概率论公理化

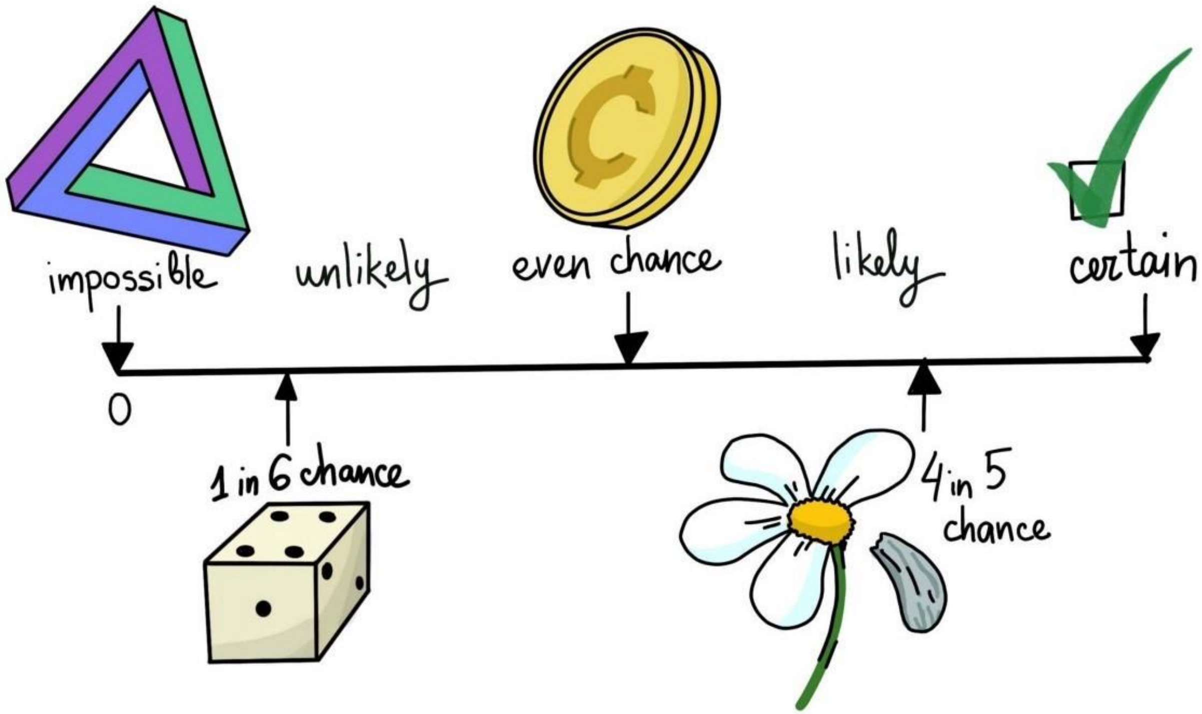
随机事件

- 随机现象总是通过随机试验来研究的.随机试验的每一种结果,称为一个随机事件.通常用大写字母

$$A,B,C,\cdots$$

来表示。

- 研究随机现象必然涉及到多个随机事件。为了掌握事件发生的规律,讨论事件之间的关系是非常必要的。
- 考虑到事件的集合内涵，往往借助于集合论的方法作为讨论事件之间关系的工具。



例： 掷一枚正方体骰子的试验,有六种可能结果,即“出现1点”,“ 出现2点”, ⋯, " 出现6点".每一种结果就看作一个事件,记作

$$A_i=\{出现i点\}\{i=1,2,\cdots,6\}$$

直观的概率 – 古典概型

虽然用频率来代替概率,但对某些事件可通过直观分析,精确地求得事件发生的概率,如：口袋中有10个大小相同的球,其中有3个红球。从中任取一个,取得红球的概率显然是$\frac{3}{10}$.

定理1：在古典概型中，若总的基本事件数为n，而事件A包含的基本事件数为m，那么A的概率可以表示为

$$P(A)=\frac{m}{n},$$

这个定义称为概率的古典定义。

例: 同时抛掷3枚硬币,求出现事件"恰有1枚正面向上“的概率.

解: 设A = {恰有1枚正面向上},

则，试验中等可能的基本事件数共有8个,即:

{正,正,正},{正,正,反},{正,反,正}{正,反,反},

{反,正,正},{反,正,反},{反,反,正}{反,反,反}.

而事件A包含3个基本事件 :

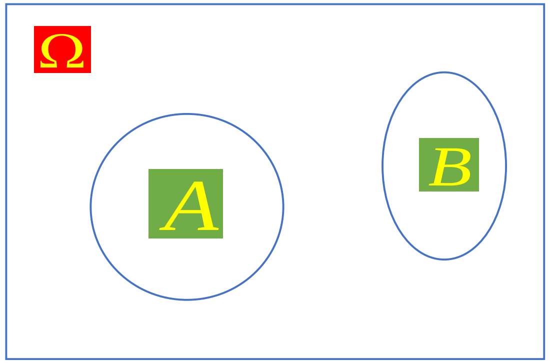
{ 正,反,反},{反,正,反},{反,反,正},

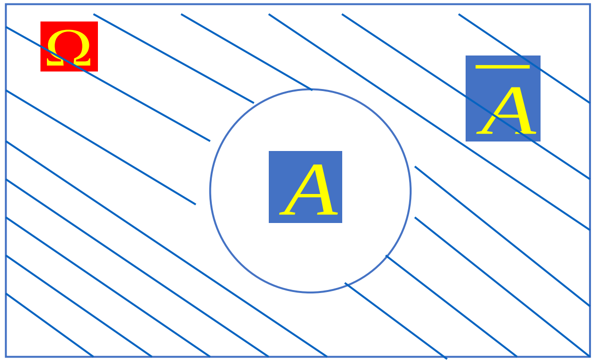
故

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

随机事件的关系与概率计算

1. 包含（$A \subset B$ 或 $B \subset A$）
2. 相等（$A \subset B$ 且 $B \subset A$）
3. 事件的和（$A \cup B$ 或 $A + B$）
4. 事件的积（$A \cap B$ 或 AB）
5. 互斥事件（$AB = \varnothing$）
6. 对立事件（$AB = \varnothing$ 且 $A + B = \Omega$）





例：有10件产品，其中7件是正品，无放回地抽取3件，求：

(1) 这3件全是正品概率;

(2)这3件恰有2件是正品的概率

解：设$A = \{全是正品\}, B = \{恰好2件是正品\}$,

从10件中任取3件共有C_{10}^3种等可能的基本事件.

(1)3件便正品的取法有C_{10}^3种, 故

$$P(A) = \frac{C_7^3}{C_{10}^3} = \frac{35}{120} = \frac{7}{24}$$

(2)3件中恰有2件正品的概率

$$P(B) = \frac{C_7^2 C_3^1}{C_{10}^3} = \frac{63}{120} = \frac{21}{40}$$

1. 互斥事件的概率加法公式

定理2： 如果A与B是互斥事件，那么A与B的概率等于它们概率的和，即：

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

推论1： 若$A_1, A_2, \cdots, A_n$两两互斥，则

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

推论2： 事件A的概率等于1减去它的对立事件的概率，即

$$P(A)=1-P(A)$$

2. 任意事件的概率加法公式

定理3： 若A与B是任意事件。那么A与B的和事件的概率等于A与B的概率之和减去A与B的积的概率，即

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

例： 从1,2,⋯,100的正整数中任取一个，求下列事件的概率：

(1) 被抽取的数能被5或21整除；

(2) 被抽取的数能被5或6整除.

解： 记事件被抽取的数，

$A=\{$能被5整除$\}$, $B=\{$能被21的整除$\}$, $C=\{$能被6整除$\}$.

(1)因A,B为互斥事件，故

$$P(A+B)=P(A)+P(B)=\frac{20}{100}+\frac{4}{100}=\frac{6}{25}$$

(2)因30,60,90既被5整除又能被6整除，所以A,C不是互斥事件，故

$$P(A+C)=P(A)+P(C)-P(AC)=\frac{20}{100}+\frac{16}{100}-\frac{3}{100}=\frac{33}{100}$$

概率论的公理化

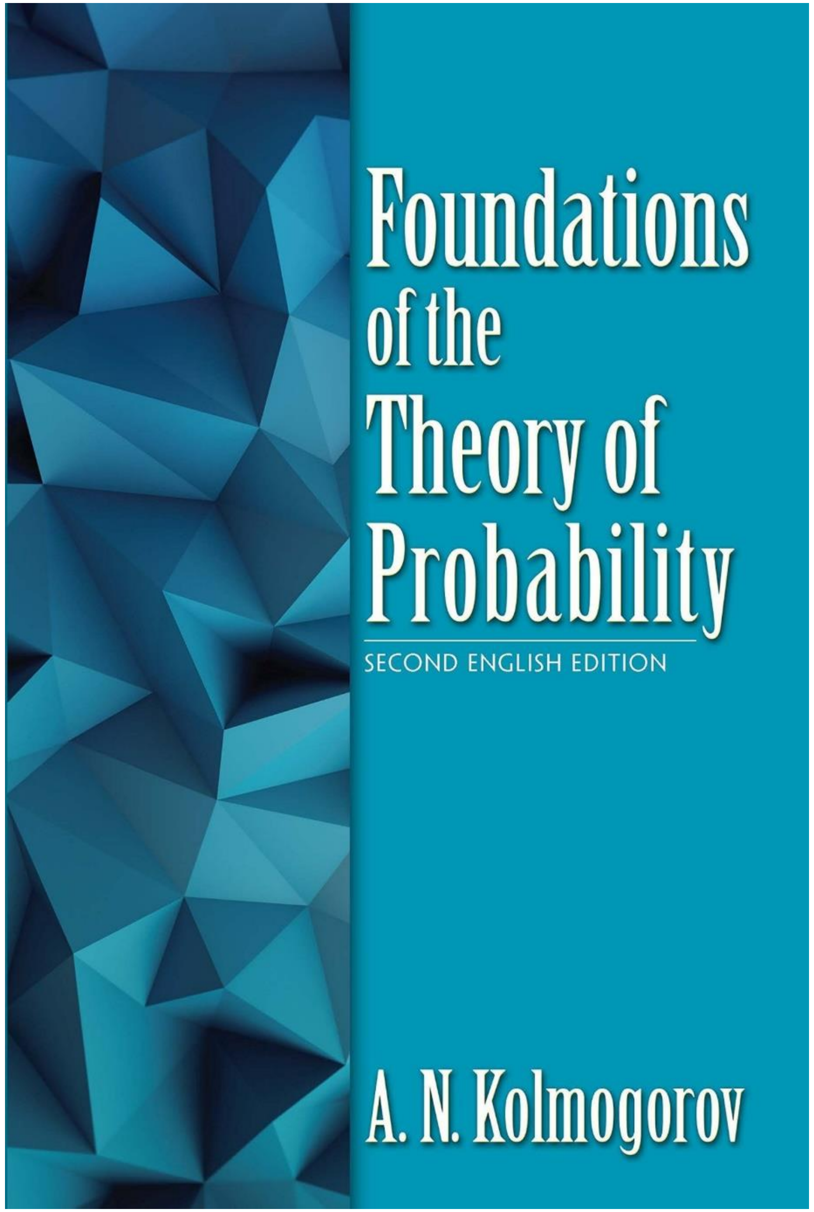
20世纪初测度和积分理论成熟，为概率公理体系的建立奠定了基础，柯尔莫哥洛夫《概率论基础》，1933年。

➢ 设随机实验**E**的样本空间为$\boldsymbol{\Omega}$。若按照某种方法，对**E**的每一事件A赋于一个实数P(A)，且满足以下公理：

1. 非负性$P(A) \geq 0$
2. 规范性$P(\Omega) = 1$
3. 可列可加性：对于两两不相容的无穷多个事件 $\boldsymbol{A_1, A_2, \cdots, A_n, \cdots}$，有：

$$\boldsymbol{P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots}$$

则称实数$\boldsymbol{P(A)}$为事件$\boldsymbol{A}$的概率.记概率空间$(\boldsymbol{\Omega}, \mathbf{E}, \mathbf{P})$.



概率论及其公理化

- 1909：Borel 尝试用测度论建立概率论基础
- 1923：Lomnicki讨论了这一思想的某些方面
- 1900s：Bohlmann 尝试了概率论的公理化
- 1917：Bernshtein 有关概率论基础构建的文章。其中，事件的集合被看做是布尔代数，并且是根据随机事件的概率大小进行定性比较。
- von Mises对概率理论的基础采用了另一种方法，将随机事件的概率与某种理想实验的结果关联起来，并需要假设该结果的频率极限的存在性。

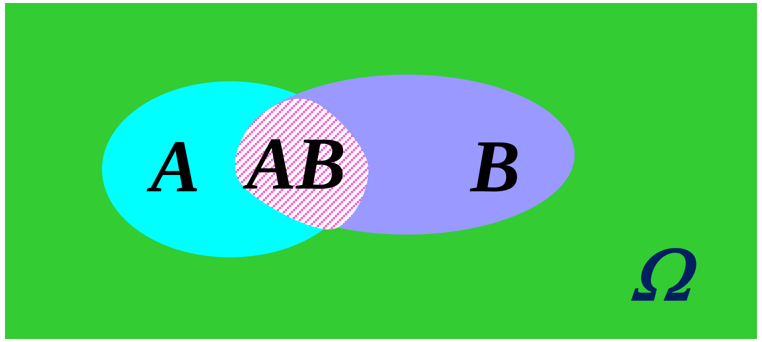
➢ 更多内容参考：Shiryaev院士在1989年，为他的老师Kolmogorov写的长文《Kolmogorov：life and creative activities》。**中译名为《老师柯尔莫哥洛夫的生平和工作》**

进阶：条件概率

若 Ω 是全集， A, B 是其中的事件，P表示事件发生的概率，则

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

为事件B发生后A发生的概率。



例：假设我国人口中能活75岁的概率为0.8活到100岁以上的概率为0.2。问：有一个已经活到75岁的老人,问能活100岁以上的概率是多少？

解：设A={活到100岁},B={活到75岁},则$P(B)$=0.8。

由于活到100岁的人必活到75岁,所以

$$AB=A,$$

因而有$P(AB)=P(A)=0.2$.

故有：

$$P(A|B) = \frac{P(AB)}{P(B)} = \frac{0.2}{0.8} = 0.25$$

A. 概率的乘法公式

由条件概率的定义$P(A|B)=\dfrac{P(AB)}{P(B)}$立即可推出：

$$P(AB)=P(B)P(A|B),P(B)>0$$

进一步推广：设$A_1,A_2,\cdots,A_n$ 为 n 个事件,$n\geq 2$,且 $P(A_1A_2\cdots A_{n-1})>0$，则有

$$\boxed{\begin{aligned}P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)&\\ \cdots P(A_n|A_1A_2\cdots A_{n-1})&\end{aligned}}$$

例：在100个零件中有4个次品,从中接连抽取两次,每次取一个,无放回地抽取,求下列事件的概率:

(1) 第二次才取正品; (2) 两次都取到正品; (3) 两次中恰取到一个正品.

解：设A_i={第i次取到正品},(i=1,2).

(1)设$A=\{$第二次才取正品$\}$,那么A表示第一次取到次品与第二次取到正品同时发生,则

$$P(A)=P(\overline{A_1}A_2)=P(\overline{A_1})P(A_2|\overline{A_1})=\frac{4}{100}\times\frac{96}{99}=0.0388$$

(2)设$B=\{$两次都取到正品$\}$,那么B表示第一次取到正品与第二次取到正品同时发生,则

$$P(B)=P(A_1A_2)=P(A_1)P(A_2|A_1)=\frac{96}{100}\times\frac{95}{99}=0.9212$$

(3)设$C=\{$两次中恰有一个正品$\}$,那么C表示"第一次取到正品且第二次取到次品"发生,则

$$P(C)=P\left(A_1\overline{A_2}+\overline{A_1}A_2\right)=P\left(A_1\overline{A_2}\right)+P\left(\overline{A_1}A_2\right)=P(A_1)P\left(\overline{A_2}|A_1\right)+P\left(\overline{A_1}\right)P\left(A_2|\overline{A_1}\right)$$
$$=\frac{96}{100}\times\frac{4}{99}+\frac{4}{100}\times\frac{96}{99}=0.0776$$

B. 全概率公式

定理：如果事件$A_1,A_2,\cdots,A_n$满足：

① $A_1,A_2,\cdots,A_n$两两互斥，且

$$P(A)>0,(i=1,2,\cdots,n)$$

② $A_1+A_2+\cdots A_n=\Omega$

则，对于任意事件B，有

$$P(B)=P(A_1)P(B|A_1)+P(A_2)P(B|A_2)+\cdots P(A_n)P(B|A_n),$$

即：

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

例：某高射炮向敌机发射一枚炮弹,已知该炮弹能击中敌机的发动机,机舱及其他部位的概率分别为0.15,0.1,0.4,又知击中上述各部位而使敌机坠毁的概率为0.9,0.85,0.55,求该高射炮发射一枚炮弹而使敌机坠毁的概率.

解：设A_1, A_2, A_3, A_4分别表示炮弹击中发动机,机舱,其他部位及击不中事件，B表示敌机坠毁的事件。则可知

$$P(A_1)=0.15,P(A_2)=0.1,P(A_3)=0.4,\quad P(A_4)=0.35$$

又因为

$$P(B|A_1)=0.9,P(B|A_2)=0.85,P(B|A_3)=0.55,P(B|A_4)=0$$

由全概率公式，可得

$$P(B)=\sum_{i=1}^{4}P(A_i)P(B|A_i)$$

$$=0.15\times0.9+0.1\times0.85+0.4\times0.55+0.35\times0$$

$$=0.135+0.085+0.220+0=0.440$$

概率分析

在我方某前沿防守地域，敌人以一个炮排（含两门火炮）为单位对我方进行干扰和破坏．为躲避我方打击，敌方对其阵地进行了伪装并经常变换射击地点。

经过长期观察发现，我方指挥所对敌方目标的指示有50%是准确的，而我方火力单位，在指示正确时，有1/3的射击效果能毁伤敌人一门火炮，有1/6的射击效果能全部消灭敌人

现在希望能用某种方式把我方将要对敌人实施的20次打击结果显现出来，确定有效射击的比率及毁伤敌方火炮的平均值

- 理论计算(1/3)：$1\times\frac{1}{2}\times\frac{1}{3}+2\times\frac{1}{2}\times\frac{1}{6}=\frac{1}{3}$

统计模拟

分析：需要模拟出以下两件事

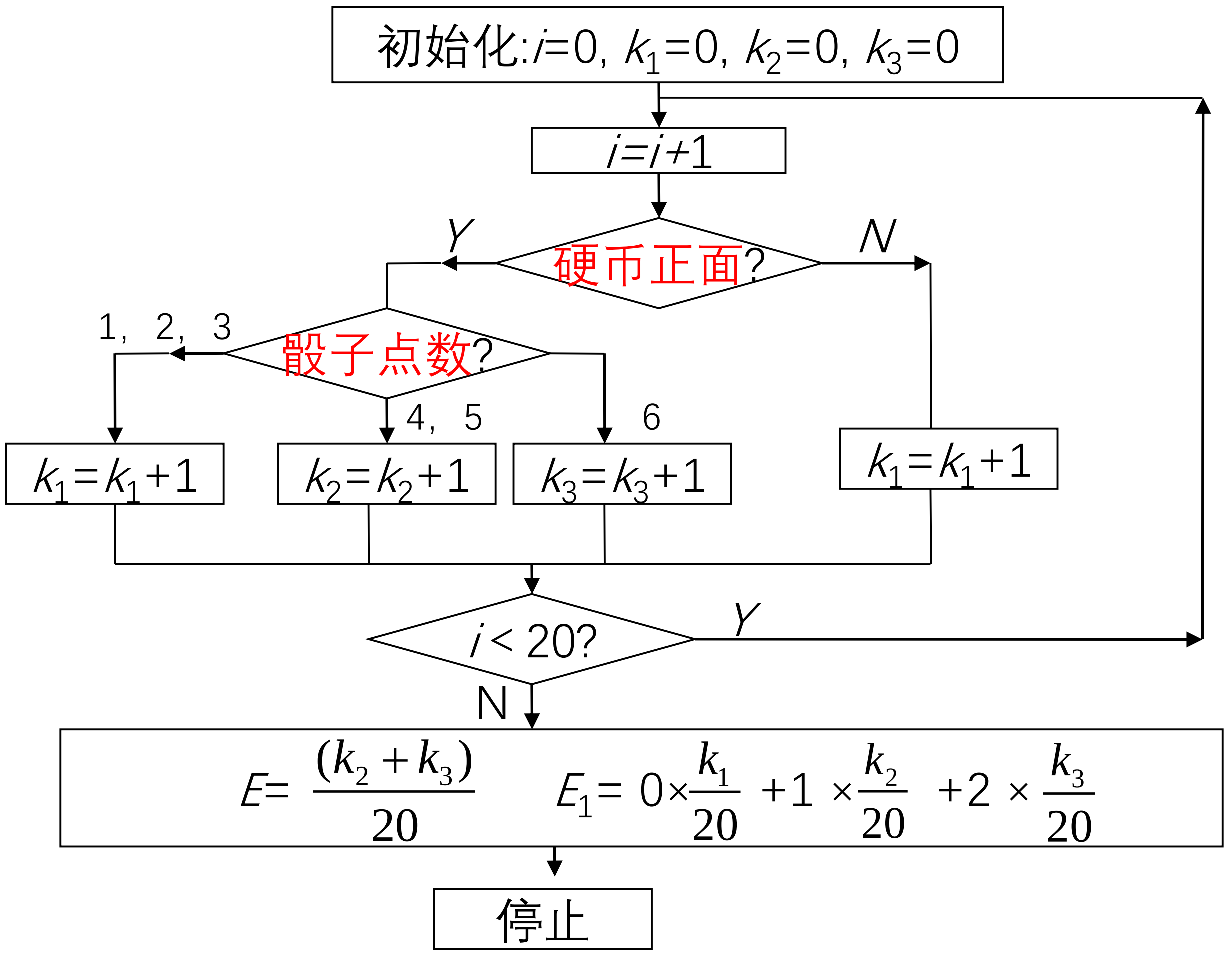
[1] **观察所对目标的指示正确与否？**模拟试验有两种结果，每种结果出现的概率都是1/2．因此，**可用投掷1枚硬币的方式予以确定**，当硬币出现正面时为指示正确，反之不正确．

[2] **当指示正确时，我方火力单位的射击结果情况：**模拟试验有三种结果：毁伤1门火炮的可能性为1/3(即2/6)，毁伤两门的可能性为1/6，没能毁伤敌火炮的可能性为1/2(即3/6)．这时**可用投掷骰子的方法来确定**：

- 如果出现的是1、2、3点：则认为没能击中敌人;
- 如果出现的是4、5点：则认为毁伤敌人一门火炮;
- 若出现的是6点：则认为毁伤敌人两门火炮．

模拟算法参考流程

- i：要模拟的打击次数；
- k_1：没击中敌人火炮的射击总数；
- k_2：击中敌人一门火炮的射击总数；
- k_3：击中敌人两门火炮的射击总数 .
- E：有效射击比率；
- E_1：20次射击平均每次毁伤敌人的火炮数 .



模拟结果

试验序号	投硬币结　果	指示正确	指　示不正确	掷骰子结　果	消灭敌人火炮数		
					0	1	2
1	正	V		4		V	
2	正	V		4		V	
3	反		V		V		
4	正	V		1	V		
5	正	V		2	V		
6	反		V		V		
7	正	V		3	V		
8	正	V		6			V
9	反		V		V		
10	反		V		V		

试验序号	投硬币结　果	指示正确	指　示不正确	掷骰子结　果	消灭敌人火炮数		
					0	1	2
11	正	V		2	V		
12	反		V		V		
13	正	V		3	V		
14	反		V		V		
15	正	V		6			V
16	正	V		4		V	
17	正	V		2	V		
18	正	V		4		V	
19	反		V		V		
20	正	V		6			V

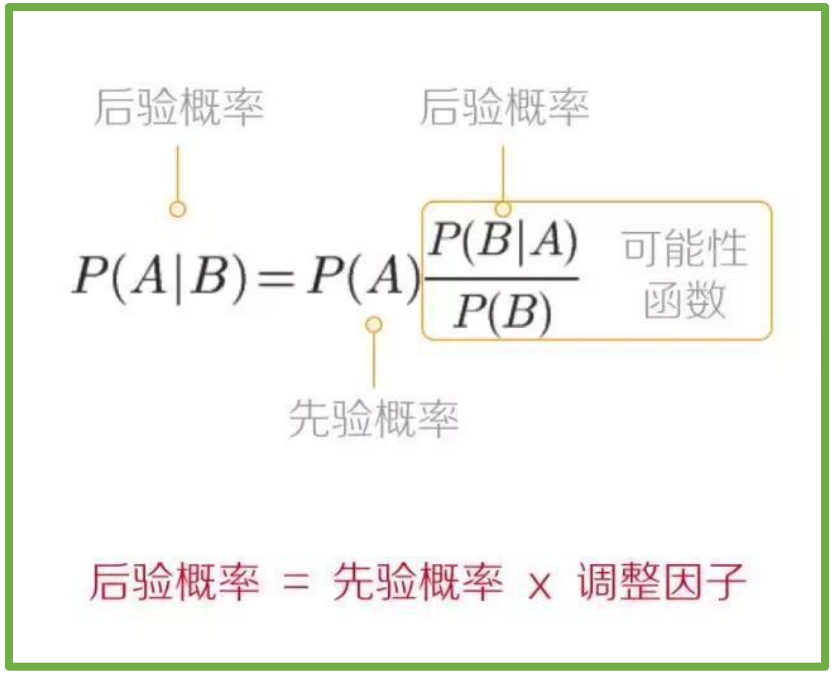
从以上模拟结果可计算出： $E=7/20=0.35$ $E_1=0\times\frac{13}{20}+1\times\frac{4}{20}+2\times\frac{3}{20}=0.5$

C. 贝叶斯公式

容易根据概率乘法公式证明:

$$P(AB) = P(B)P(A|B), P(B) > 0$$

$$P(AB) = P(A)P(B|A), P(A) > 0$$



应用举例:

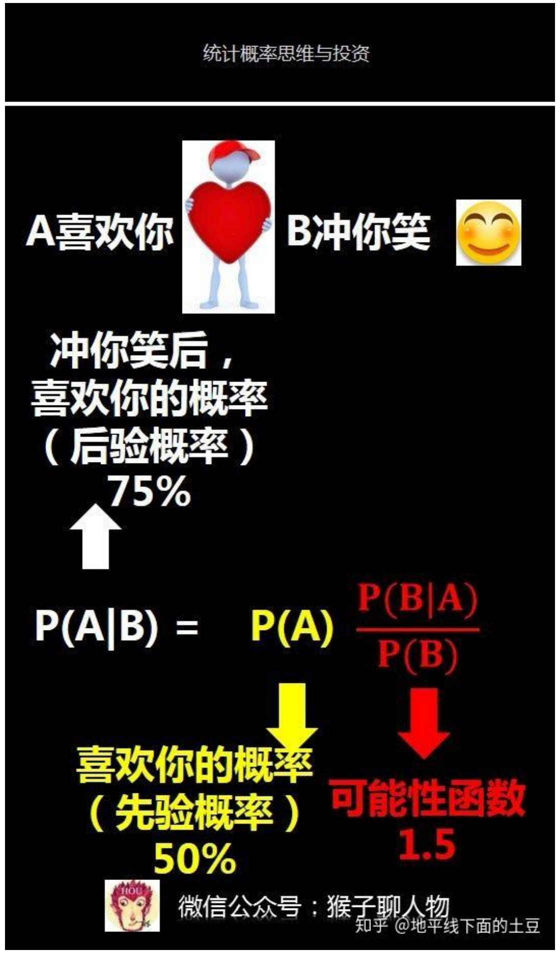
已知:

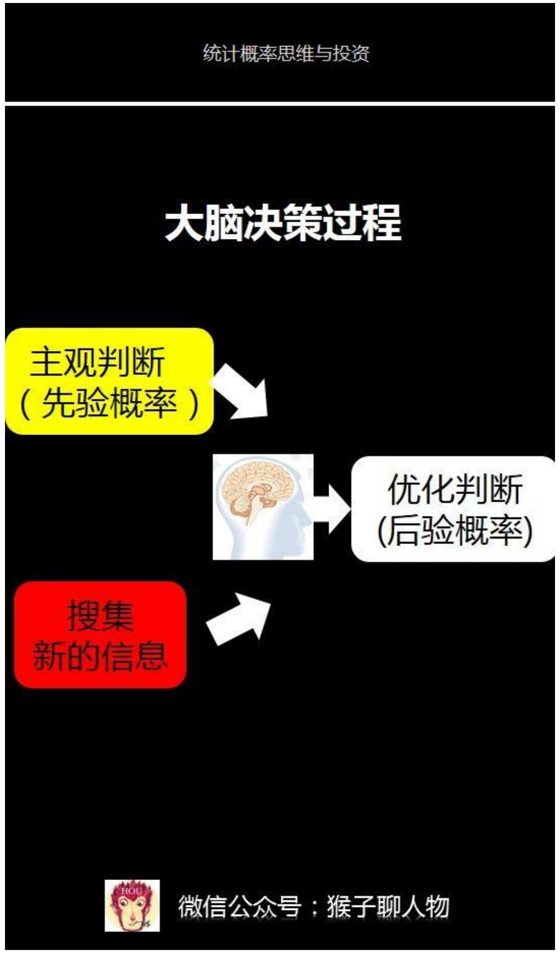
- 病人患有脑膜炎(事件A)的先验概率是1/50000 ： P(A)
- 病人患有颈部僵硬(事件B)的先验概率是1/20 ： P(B)
- 患脑膜炎的人有50%会发生颈部僵硬 ： P(B|A)

问：如果病人患有颈部僵硬，那么他患有脑膜炎的概率有多大?

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}=\frac{0.5\times 1/50000}{1/20}=0.0002$$

一个直观理解 (本页资料来源于网络)





阅读资料:

"怎样用数学找到一颗丢失的氢弹？"——贝叶斯定理在搜索失踪物品方面的几个小故事

https://blog.csdn.net/FnqTyr45/article/details/86635384

2. 随机变量 & 随机过程

2.1 随机变量

- 设E是随机试验,**样本空间Ω**。如果对于每一个基本事件$\omega\in\Omega$, 都有**唯一**的实数$X(\omega)$与之**对应**,则称$X(\omega)$为Ω上的一个随机变量。
- 一般用大写字母X,Y,Z表示。
- 取不同值的可能性大小是由相应的随机事件发生概率的大小决定的.

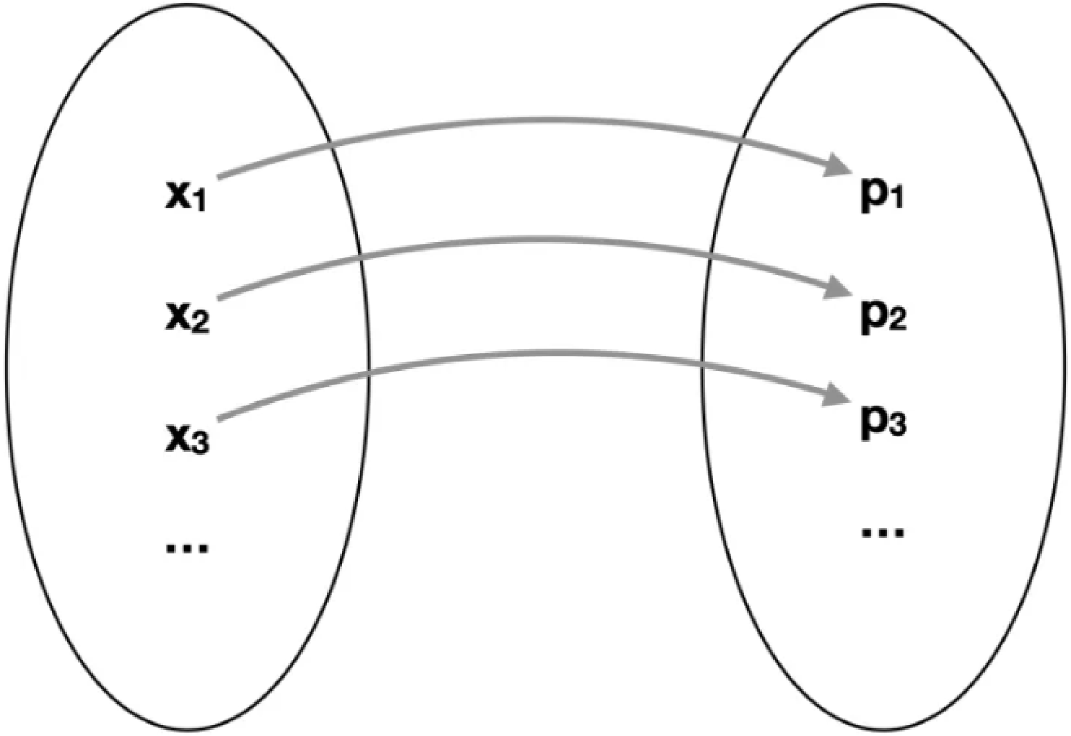
例1：设某人射击,其每次击中目标的可能性为0.8,现在他连续射击30次,则该人击中目标的次数就是一个随机变量X,并且X所有可能性的取值0,1,2,⋯,30.

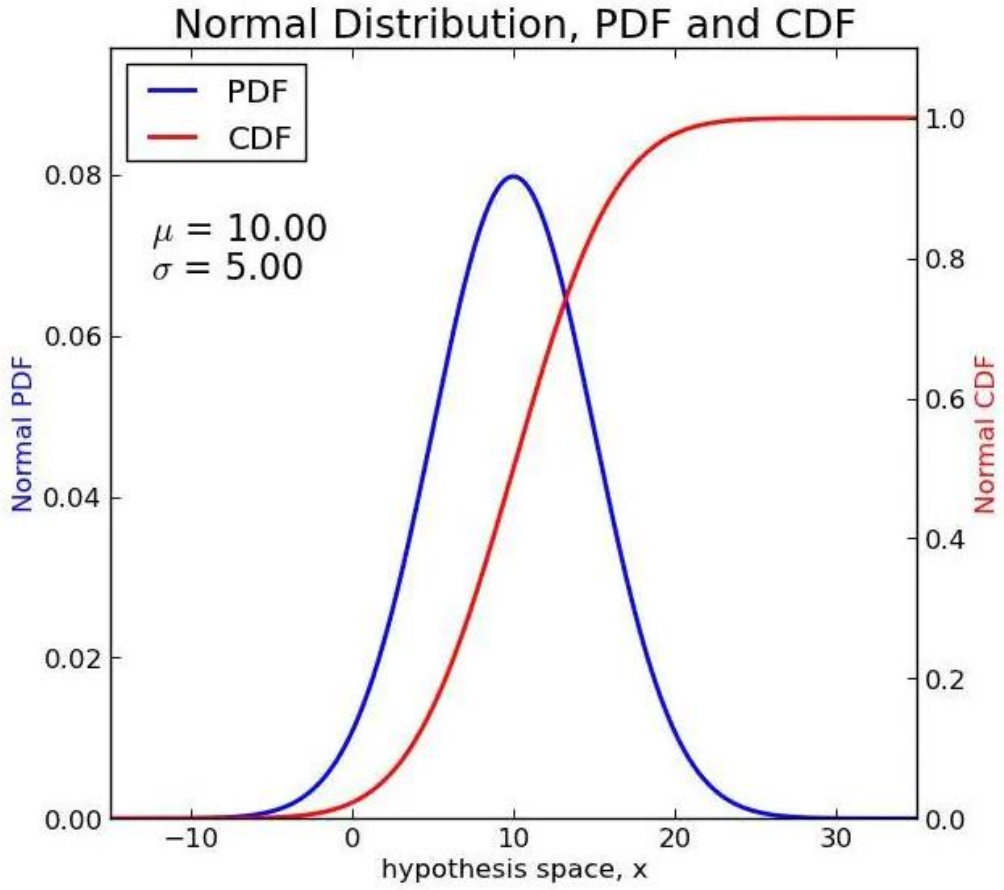
例2：某汽车站每10min一班车,有位乘客事先并不知道汽车到达车站的时间,并且他在任一时刻到达车站都是可能的,那么,他等候汽车的时间就是一个随机变量X,并且该变量的取值范围为一个区间[0,10].

随机变量的**二要素**:

（1）取值情况；　　（2）**取值的概率分布**（后续再展开）

PDF & CDF





事实上：(**连续)随机变量 X 的概率分布一般**用概率密度函数$p(x)$ 描述是方便的：$\int_{-\infty}^{\infty} p(x)dx = 1$.

Probability Distribution Function(PDF→CDF)

- 概率分布函数体现了概率统计的核心思维方式，蕴含着一个随机变量取值的概率规律：

对概率空间 $(\Omega, \mathcal{F}, P)$，其中P是一个度量，定义

$$F_X(a) = P(X \le a), \quad \forall a \in \mathbb{R},$$

称为X的分布函数 (distribution function) 或累积分布函数 (CDF) 它有如下性质：

右连续

(1) 单调不减且右连续；

阶梯是这样的.

(2) $F_X(-\infty) = 0$, $F_X(\infty) = 1$.

反之，满足(1)、(2)的函数也可称为一个分布函数.

- 常见的随机变量有两大类：
  - 离散型：$F_x(a)$取值为有限个或无穷可列个，如投骰子(1-6)
  - 连续型：$F_x(a)$取值为某个区间或整个实数域R

随机变量的数字特征

1. 数学期望

- 离散型: $EX = \sum_{k} x_k p_k$
- 连续型: $EX = \int_{-\infty}^{+\infty} xf(x)dx$

2. 方差

$E(X-EX)^2$为X的方差,记为

$$DX = E(X-EX)^2$$

3. 协方差

$$\mathrm{Cov}(X,Y) = E\{[X-E(X)][Y-E(Y)]\};$$

设 X,Y 为离散型随机变量 p_{ij}为联合分布列则

$$\mathrm{Cov}(X,Y) = \sum_i \sum_j [x_i - E(X)][y_j - E(Y)]p_{ij}.$$

设 X,Y 为连续型随机变量 $f(x,y)$ 为联合概率密度则

$$\mathrm{Cov}(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(X)][y-E(Y)]f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

4. 相关系数

$$\rho = \rho_{XY} = \frac{\mathrm{cov}(X,Y)}{\sqrt{DX}\sqrt{DY}}$$

称为X与Y的**相关系数**.

大数定律

定理1（切比雪夫大数定律）

设 $X_1, X_2, \ldots$是相互独立的随机变量序列，它们都有有限的方差，并且方差有共同的上界，即 $DX_i \leqslant K$，$i=1, 2, \ldots$，

则对任意的 $\varepsilon > 0$，

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} EX_i| < \varepsilon\} = 1$$

切比雪夫大数定律表明，独立随机变量序列$\{X_n\}$，如果方差有共同的上界，则

$\frac{1}{n}\sum_{i=1}^{n} X_i$ 与其数学期望 $\frac{1}{n}\sum_{i=1}^{n} E(X_i)$ 偏差很小的概率接近于1.

即当n充分大时，$\frac{1}{n}\sum_{i=1}^{n} X_i$ 差不多不再是随机的了，取值接近于其数学期望的概率接近于1.

切比雪夫大数定律给出了平均值稳定性的科学描述

- 证明切比雪夫大数定律的主要数学工具是切比雪夫不等式:

设随机变量X有期望EX和方差DX，则对于$\forall\varepsilon > 0$，有

$$P\{|X - EX| < \varepsilon\} \geq 1 - \frac{DX}{\varepsilon^2}$$

(弱)大数定律： more is better!

1. 切比雪夫大数定律
2. 伯努利大数定律
3. 辛钦大数定律

大数定律以严格的数学形式表达了随机现象最根本的性质之一：

平均结果(频率)的稳定性

它是随机现象统计规律的具体表现.

大数定律在理论和实际中都有广泛的应用.

下面给出的独立同分布下的大数定律，不要求随机变量的方差存在。

辛钦

定理4（辛钦大数定律）

设随机变量序列$X_1, X_2, \dots$独立同分布，具有有限的数学期$EX_i = \mu$, $i=1,2,\dots$，则对任给 $\varepsilon>0$，

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$$

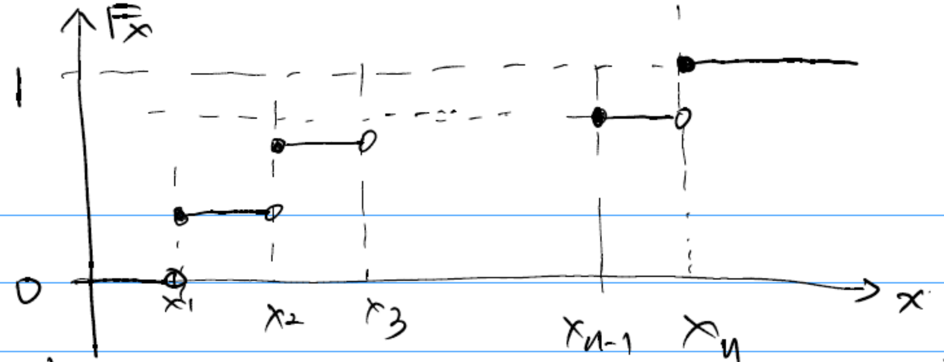
$$\boxed{\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \mu.}$$

离散型分布函数

若X只有有限(可列)的值 $x_1 < x_2 < \cdots < x_n$ (可→∞)，则分布函数是

$$F_X(x) = \begin{cases} 0, & -\infty < x < x_1, \\ \sum_{j=1}^{i} P(x_j), & x_i \le x < x_{i+1}, \\ 1, & x_n \le x < +\infty \end{cases}$$

← 累加 $P(x_j)$ 是一个 "原子事件 x_j" 概率



注意

在离散下，概率密度为：$\{P(x_i)\}$，$i=1,2,\cdots,n$. 注意 $P(x_i)$ 是上面每个台阶的高度.

一年中发生枪击案的次数	0	1	2	3	4	5	6	7
年　数	4	10	7	5	4	0	0	1

实用表达:分布列

二项分布（Binomial Distribution）

伯努利(Bernoulli)试验：

- 在一次试验中，事件A出现（如抛掷一枚硬币正面朝上）的概率为μ，不出现（硬币反面朝上）的概率为$1-\mu$。若用变量X表示事件A出现的次数，则X的取值为0或1，相应的分布（也叫两点分布$X\sim B(1,p)$）：

$$\cancel{p(x)=\mu^x(1-\mu)^{1-x}}$$

- 由定义，计算可知：

$$\mathrm{E}[X]=1\cdot p+0\cdot(1-p)=p$$

$$var(X)=(1-p)^2\cdot p+(0-p)^2\cdot(1-p)$$

$$=p(1-p).$$

二项分布

- 在n次伯努利试验中，若以变量X表示事件A出现的次数，则X的取值为$\{0,\cdots,n\}$,其相应的分布（记作$X\sim B(n,p)$）为

$$\mathrm{P}(\mathrm{X}=\mathrm{k})=\binom{n}{k}\mu^{\mathrm{k}}(1-\mu)^{n-k},$$

$$k=0,1,2,\cdots,n.$$

- 由定义，计算可知：

$$\mathrm{E}[X]=np,var(X)=np(1-p).$$

- API: binocdf/binopdf

```
>> p = binopdf(1, 100, 0.01)
```

泊松(Poisson)分布

若随机变量X满足

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

则称X服从参数为λ的泊松分布,记为$X\sim P(\lambda)$。

泊松分布是一种常见的分布,如下随机变量都服从泊松分布:

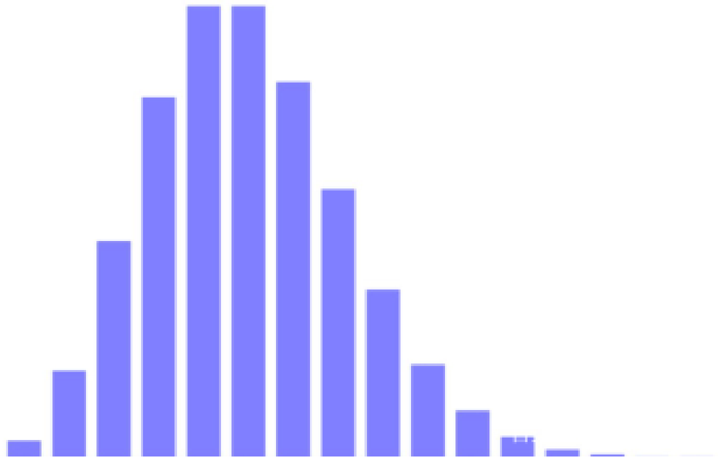
1. 在单位时间内来到电话交换局的电话呼唤次数
2. 一页书上印刷的错误数
3. 到机场降落的飞机数
4. 单位长度布匹上出现的疵点数

- 泊松分布的期望和方差均是λ!
- 例:已知$X\sim P(0.5)$,则所求事件的概率为:

$$P(X\geq 1)=1-P(X<1)=1-P(X=0)=1-e^{-\frac{1}{2}}$$

- 泊松分布(Poisson, poisspdf) :

```
>> bar(poisspdf(0:15, 5));
>> t = 0:15; plot(t+1, poisspdf(t, 5), 'k--');
```



Degree Distribution: Poisson

$$\boxed{P(k) = \frac{\mu^k}{k!} e^{-\mu}}$$

The probability of a node connecting to k nodes among other $N-1$ nodes (but not connected to the rest $N-1-k$ nodes) is given by

$$P(k|N) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

Let $\mu = p(N-1)$, which is $\langle k \rangle$ as shown before.

Let $N \to \infty$ while keeping $\mu = \langle k \rangle$ unchanged (as an expectation value). Then

$$P(k) = \lim_{N\to\infty} P(k|N) = \lim_{N\to\infty} \frac{(N-1)!}{k!\,(N-1-k)!} \left(\frac{\mu}{N-1}\right)^k \left(1 - \frac{\mu}{N-1}\right)^{N-1-k}$$

$$= \lim_{N\to\infty} \frac{(N-1)[(N-1)-1]\cdots[(N-1)^k - k + 1]}{(N-1)^k} \frac{\mu^k}{k!} \left(1 - \frac{\mu}{N-1}\right)^{N-1} \left(1 - \frac{\mu}{N-1}\right)^{-k}$$

$$= 1 \cdot \frac{\mu^k}{k!} \cdot e^{-\mu} \cdot 1$$

$$\lim_{n\to\infty} C_n^k p^k (1-p)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}$$

该定理说明：泊松分布是二项分布当$np \to \lambda$, $(n \to \infty)$极限情况下的逼近。
因此，当n很大且p很小时，可以用泊松分布作二项分布的近似计算！

连续型分布函数$f(x)$: 可微分→ PDF

定理2.1：设 $F: [-\infty, \infty] \longmapsto [0,1]$，$F(-\infty)=0$，$F(\infty)=1$，且F单调不减，右连续，则存在 $(\Omega, \mathcal{F}, P)$ 及其上的随机变量X，使F是其分布函数：$F = F_X$.

注意正如其名，F_X 是一种积累，且显然有

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a)$$
$$= F_X(b) - F_X(a).$$

由可加性和 F_X 性质，形式上可看作是一个积分：

$$P(a < X \leq b) = \int_a^b f(x)\,dx = F_X(x)\Big|_a^b = F_X(b) - F_X(a)$$

因此有

$$\boxed{f(x) = F_X'(x).}$$

称为X的概率密度函数(probability density function, PDF)。它直观上更接近我们之前提到的"原子事件"，但对连续分布，其中一个点是零概率事件，而将一个区间收缩到一点的极限，就是PDF在该点的值。而一个连续区间对应事件的概率，只需对 $f(x)$ 积分即可，因此非常方便。在连续分布中，我们更多采用PDF描述。

正态分布（Normal Distribution）

1. 正态(normpdf,randn)分布

$$X \sim \mathcal{N}(\mu, \sigma^2) \qquad p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Matlab 提供的生成正态概率密度函数和累积分布函数的命令分别是：

normpdf(X, a, b)：给出概率密度函数在 X 各个点上的值；

normcdf(X, a, b)：给出累积分布函数在 X 各个点上的值，

其中 a 和 b 分别是分布的期望和标准差，如果缺省该两个参数，则是 $a = 0, b = 1$ 的情形，即为标准正态分布。

注意　在 Matlab 的正态函数命令中，使用的不是方差而是标准差。

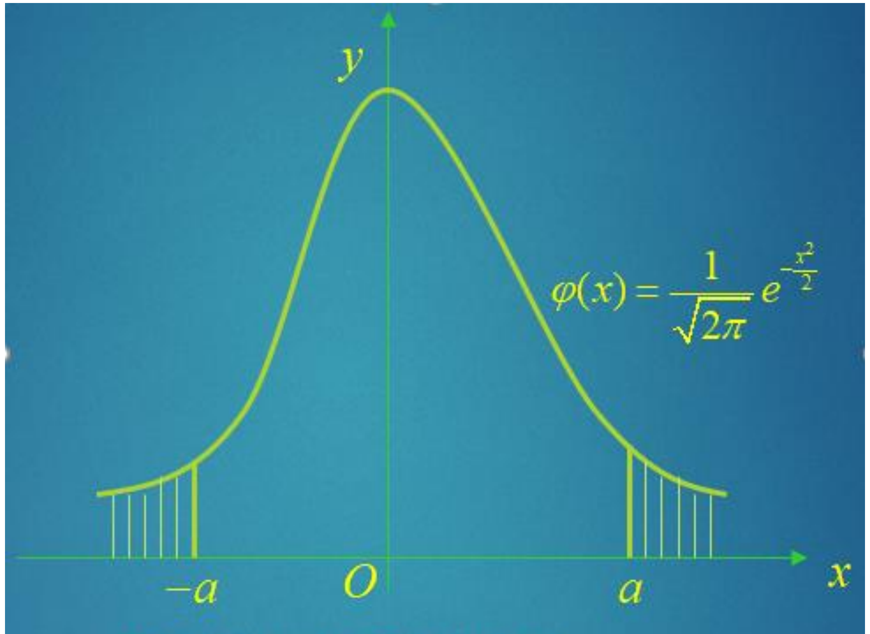


图7-11 标准正态曲线

2. 对数正态(lognpdf, lognrnd)分布

$$f(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \mathrm{e}^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}}, & x > 0, \\ 0, & x \leqslant 0。 \end{cases}$$

中心极限定理

设$X_1, X_2, \cdots, X_n, \cdots$独立同分布，

$E(X_i)=\mu, Var(X_i)=\sigma^2$,则对任意实数$x$,

$$\lim_{n\to\infty} P\left(\frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

因此当n充分大时 $\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2)$.

$$\frac{1}{n}\sum_{i=1}^{n} X_i \overset{\text{近似}}{\sim} N(\mu, \frac{\sigma^2}{n})$$

推论(棣莫弗—拉普拉斯中心极限定理)：

设n_A为n重贝努里试验中A发生的次数，$P(A)=p(0<p<1)$,则对任何实数x，有：

$$\lim_{n\to+\infty} P\left(\frac{n_A - np}{\sqrt{np(1-p)}} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x).$$

即，$B(n,p) \overset{\text{近似}}{\sim} N(np, np(1-p))$，当$n$充分大时.

证明：令$X_i = \begin{cases} 1, & \text{第}i\text{次试验}A\text{发生}, \\ 0, & \text{第}i\text{次试验}A\text{不发生}, \end{cases}$ $i=1,2,\cdots,n.$

$X_1, X_2, \ldots X_n$独立同$\sim B(1,p)$. $n_A = \sum_{i=1}^{n} X_i$,由定理可得.

指数(Exponential)分布

当人们考察相继发生事件的时间间隔，或者事件的存续时间，往往发现这些时间的长度是随机的，指数分布常被用来刻画它们。指数概率密度函数为

$$f(x) = \frac{1}{a}\mathrm{e}^{-\frac{x}{a}}, \quad x \geqslant 0, \tag{4.6}$$

其中参数 $a > 0$ 是分布的期望或标准差，而其倒数 a^{-1} 则反映在单位时间内发生时间的次数，即发生率。

Matlab 提供的生成指数概率密度函数和累积分布函数的命令分别是：

exppdf(X, a)：给出概率密度函数在 X 各个点上的值；

expcdf(X, a)：给出累积分布函数在 X 各个点上的值；

exprnd(a)：生成服从参数为 a 的指数分布的随机数；

exprnd(a, [M, N])：生成由服从指数分布的随机数所组成的 $M \times N$ 矩阵 (分布参数

2.2 随机过程

- 随机变量是独立同分布的。
- **随机过程**是随机变量的集合

$$\{X_t, t \in T\}$$

指标集T可以是离散/连续的。

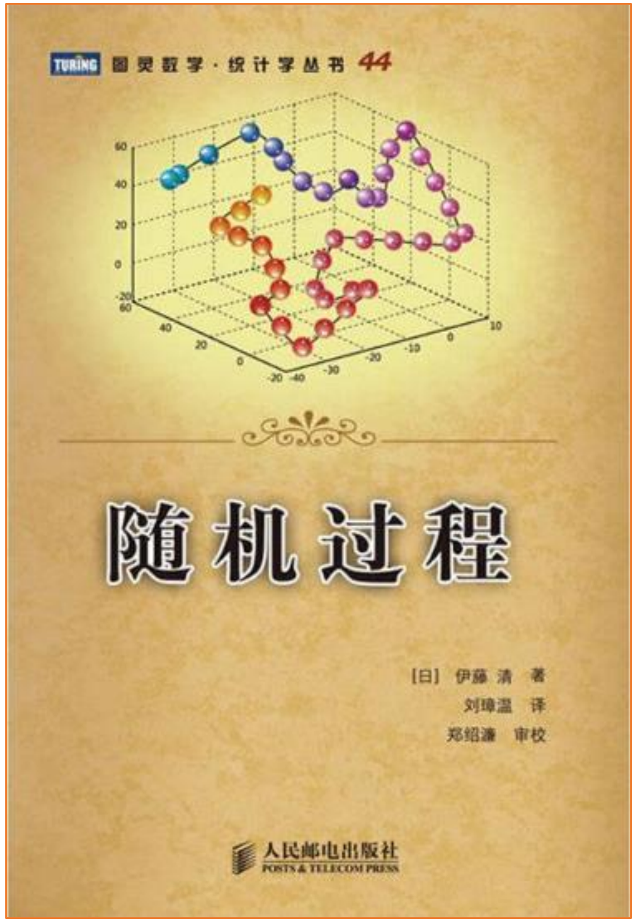
- 随机变量序列：T离散的情形。
- 常用的随机过程：
  1. Poisson过程
  2. Markov过程

例：状态集S={晴，云，雨}，指标集T={0,1,2,...}，

则称序列

晴，云，云，晴，云，雨，…

为具有离散状态空间S和离散指标集T的随机过程。

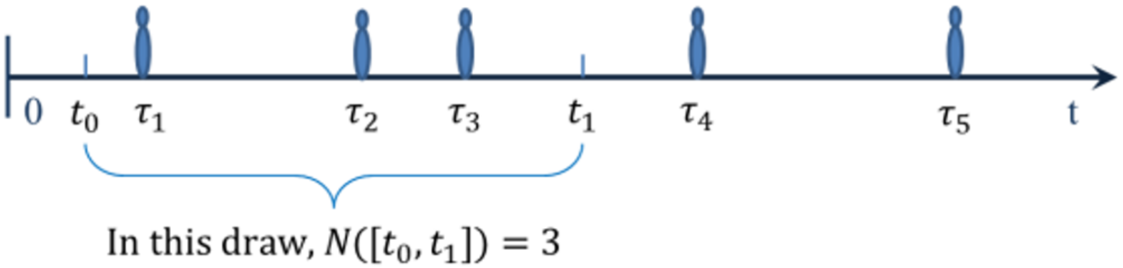


伊藤清(Kiyoshi Ito):1915-2008

- 因在概率论方面的奠基性工作被授予1987年的沃尔夫奖
- 被誉为“现代随机分析之父”：
  - 伊藤引理
  - 伊藤积分
  - 伊藤过程
- 在大一时看到过《概率论基础概念》当时毫无兴趣；在大学毕业后,一口气把它读完!

Poisson过程

假设事件在$[0,t]$上任意时刻发生。令$N(t)$表示在时间$[0,t]$上事件发生的个数。如果



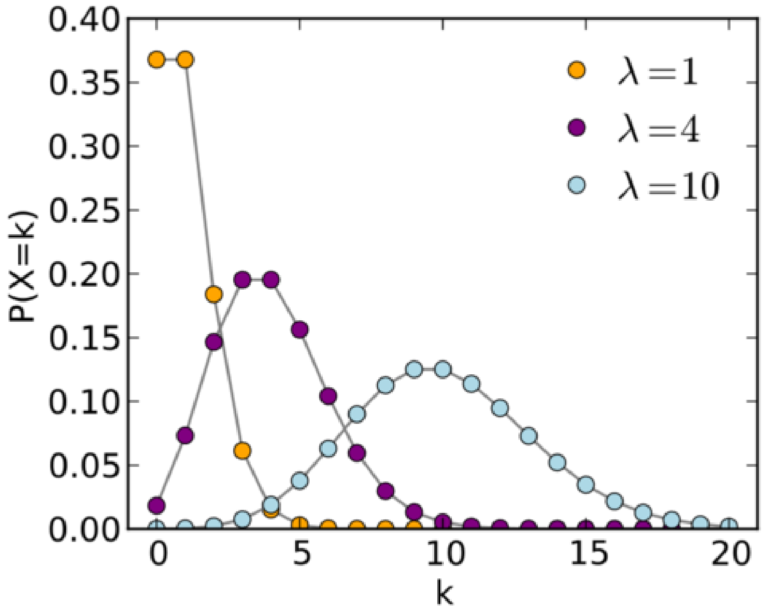
1. $N(0) = 0$
2. $N(t)$是独立增量的过程，即取任意$0 < t_1 < t_2 < \cdots < t_n$,

$$N(t_1), N(t_2) - N(t_1), \cdots, N(t_n) - N(t_{n-1})$$

相互独立;

3. $\forall\, t > 0, s \geq 0$,增量$N(s+t) - N(s)$服从参数为$\lambda t$的Poisson分布

$$P\{N(s+t) - N(s) = k\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}.$$



则称这些事件构成一个具有速率$\lambda(\lambda > 0)$的Poisson过程。

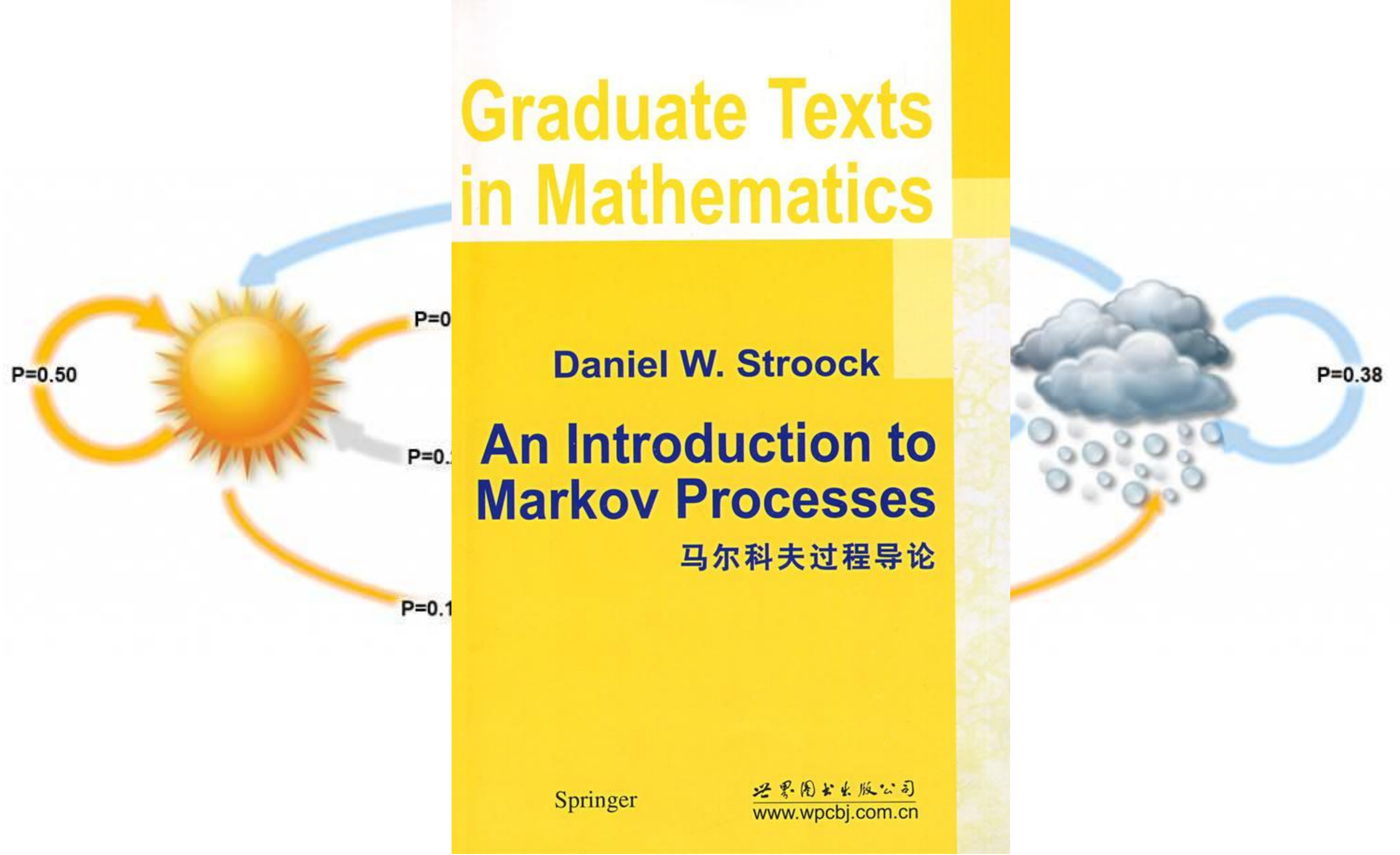
Markov过程

➢未来的演变不依赖于其以往的演变，考虑随机过程$X = \{X_t, t \in R\}$为例：

$$P(X_{t+1} = x_{t+1} \mid X_0 = x_0, \cdots, X_t = x_t) = P(X_{t+1} = x_{t+1} \mid X_t = x_t)$$

➢典型的Markov过程示例：

- 荷花池中一只青蛙的跳跃
- 液体中微粒所作的布朗运动
- 传染病受感染的人数
- 原子核中一自由电子的跳跃
- 人口增长过程
- 每天的股价波动



Homework 02

1. 列举5种以上常见的随机变量分布，并给出：
   1. 它们的均值、方差及其计算过程
   2. 在编程语言中，其密度函数、随机数的调用实例
2. 关于贝叶斯分类器
   1. 请简述其分类原理
   2. 重现实例demo_email-span.pdf，尽可能简短

“在很大程度上，人生最重要的问题就是概率问题。”

---- **皮埃尔•西蒙•拉普拉斯**

3. 贝叶斯(Bayes)分类实例

贝叶斯分类

设X是类标号未知的数据样本，H为某种假定(Hypothesis)，如数据样本X属于某特定的类C。对于分类(Classsify)问题，我们希望确定$P(H|X)$，即给定观测数据样本X，假定H成立的概率。

➢贝叶斯定理给出了如下计算$P(H|X)$的简单有效的方法:

$$P(H|X)=\frac{P(X|H)P(H)}{P(X)}$$

➢$P(H)$称为先验概率，或称H的先验概率

➢$P(H|X)$称为后验概率，或称条件X下H的后验概率

分类器：朴素贝叶斯(Naive Bayes)

问题描述：提取样本特征为n维特征向量：$X=\{x_1, x_2, \cdots\cdots, x_n\}$，分别属于$m$个类$C_1, C_2, \cdots, C_m$．给定一个未知的数据样本$X$（即没有类标号），需确定其类别:对未知样本X分类，也就是对每个类Ci，计算P(X|Ci)P(Ci)。若要将样本X归到类Ci，当且仅当

$$\mathbf{P}(\mathbf{C_i}|\mathbf{X}) > \mathbf{P}(\mathbf{C_j}|\mathbf{X}), 1 \le j \le m,\ j \ne i.$$

$$\mathbf{P}(\mathbf{C_i}|\mathbf{X}) = \frac{\mathbf{P}(\mathbf{X}|\mathbf{C_i})\mathbf{P}(\mathbf{C_i})}{\mathbf{P}(\mathbf{X})}$$

分析：P(X)对于所有类为常数，由Bayes公式，只需比较$\mathbf{P}(\mathbf{X}|\mathbf{C_i})\mathbf{P}(\mathbf{C_i})$，分两个情况：

1. 若Ci类的先验概率未知，则通常假定这些类是等概率：P(C1)=P(C2)=…=P(Cm)，问题转换为P(X|Ci)的最大化。（注：给定Ci时数据X的似然度，使P(X|Ci)最大的假设Ci称为最大似然假设）

2. 若P(Ci)可计算，则最大化P(X|Ci)P(Ci)：

   - 类的先验概率可以用P(Ci)=si/s计算，其中si是类Ci中的训练样本数，而s是训练样本总数。
   - 给定具有许多属性的数据集，计算P(X|Ci)的开销可能非常大。作类条件独立的**朴素假定**：

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

即属性间不存在依赖关系。此时可计算联合概率分布 $\mathbf{P}(\mathbf{X}|\mathbf{C_i}) = \prod_{k=1}^{n} \mathbf{P}(\mathbf{x_k}|\mathbf{C_i})$

Bayes分类案例：关于“打网球”的决定

通过调查问卷的方式获取了如下数据：

No.	天气	气温	湿度	风	*类别*
1	晴	热	高	无	*N*
2	晴	热	高	有	*N*
3	多云	热	高	无	*P*
4	雨	适中	高	无	*P*
5	雨	冷	正常	无	*P*
6	雨	冷	正常	有	*N*
7	多云	冷	正常	有	*P*

No.	天气	气温	湿度	风	*类别*
8	晴	适中	高	无	*N*
9	晴	冷	正常	无	*P*
10	雨	适中	正常	无	*P*
11	晴	适中	正常	有	*P*
12	多云	适中	高	有	*P*
13	多云	热	正常	无	*P*
14	雨	适中	高	有	*N*

那么，请对下面的情况做出决策：

天气	温度	湿度	有风	打网球
晴	冷	高	有	？

第一步:

No.	天气	气温	湿度	风	*类别*
1	晴	热	高	无	*N*
2	晴	热	高	有	*N*
3	多云	热	高	无	*P*
4	雨	适中	高	无	*P*
5	雨	冷	正常	无	*P*
6	雨	冷	正常	有	*N*
7	多云	冷	正常	有	*P*

No.	天气	气温	湿度	风	*类别*
8	晴	适中	高	无	*N*
9	晴	冷	正常	无	*P*
10	雨	适中	正常	无	*P*
11	晴	适中	正常	有	*P*
12	多云	适中	高	有	*P*
13	多云	热	正常	无	*P*
14	雨	适中	高	有	*N*

统计先验概率

天气		温度		湿度		有风		打网球	
P	N	P	N	P	N	P	N	P	N
晴 2/9	3/5	热 2/9	2/5	高 3/9	4/5	否 6/9	2/5	9/14	5/14
云 4/9	0/5	中 4/9	2/5	正常 6/9	1/5	是 3/9	3/5		
雨 3/9	2/5	冷 3/9	1/5						

第二步：根据先验概率(下表同前一页表)

天气 E1		温度 E2		湿度 E3		有风 E4		打网球 D	
P	N	P	N	P	N	P	N	P	N
晴 2/9	3/5	热 2/9	2/5	高 3/9	4/5	否 6/9	2/5	9/14	5/14
云 4/9	0/5	中 4/9	2/5	正常 6/9	1/5	是 3/9	3/5		
雨 3/9	2/5	冷 3/9	1/5						

做出决策：

天气	温度	湿度	有风	打网球
晴	冷	高	是	?

整理模型： $E = E_1 \cap E_2 \cap E_3 \cap E_4$

决策： $P(D = no|E) > P(D = yes|E)$
?

E为第二个表中的取值、分别计算：D = yes/no的概率

统计结果(下表同前一页表)

天气 E1		温度 E2		湿度 E3		有风 E4		打网球 D	
P	N	P	N	P	N	P	N	P	N
晴 2/9	3/5	热 2/9	2/5	高 3/9	4/5	否 6/9	2/5	9/14	5/14
云 4/9	0/5	中 4/9	2/5	正常 6/9	1/5	是 3/9	3/5		
雨 3/9	2/5	冷 3/9	1/5						

对下面的情况做出决策:

天气	温度	湿度	有风	打网球
晴	冷	高	是	?

贝叶斯公式:

$$P(D|E) = \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E)}$$

Case “yes”:

$$P(yes|E) = \frac{P(DE)}{P(E)} = \frac{P(E|D)P(D)}{P(E)}$$

$$= \frac{\boxed{P(E_1 \cap E_2 \cap E_3 \cap E_4|yes)}P(yes)}{P(E)}$$

$$= \frac{\boxed{P(E_1|yes)P(E_2|yes)P(E_3|yes)P(E_4|yes)P(yes)}}{P(E)}$$

$$P(yes|E) = \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{P(E)} = \frac{0.0053}{P(E)}$$

同理可计算:

$$P(no|E) = \frac{\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}}{P(E)} = \frac{0.0206}{P(E)}$$

已经计算出：

$$P(yes|E) = \frac{\frac{2}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{3}{9} \times \frac{9}{14}}{P(E)} = \frac{0.0053}{P(E)}$$

同理可计算：

$$P(no|E) = \frac{\frac{3}{5} \times \frac{1}{5} \times \frac{4}{5} \times \frac{3}{5} \times \frac{5}{14}}{P(E)} = \frac{0.0206}{P(E)}$$

利用公式：

$$P(yes|E) + P(no|E) = 1$$

最后得到：

$$P(yes|E) = 20.5\% \qquad P(no|E) = 79.5\%$$

或可直接决策：

$$P(yes|E) < P(no|E)$$ 不去打球