深圳北理莫斯科大学
УНИВЕРСИТЕТ МГУ-ППИ В ШЭНЬЧЖЭНЕ
SHENZHEN MSU-BIT UNIVERSITY

Математическое моделирование и исследование моделей с помощью математических программ

数学建模及数学软件的使用

Лекция № 9 (概率论)

张晔

ye.zhang@smbu.edu.cn

概率论/ Probability theory/ Теория вероятностей

概率论是研究什么的？

随机现象：不确定性与统计规律性

概率论——研究和揭示随机现象的统计规律性的科学

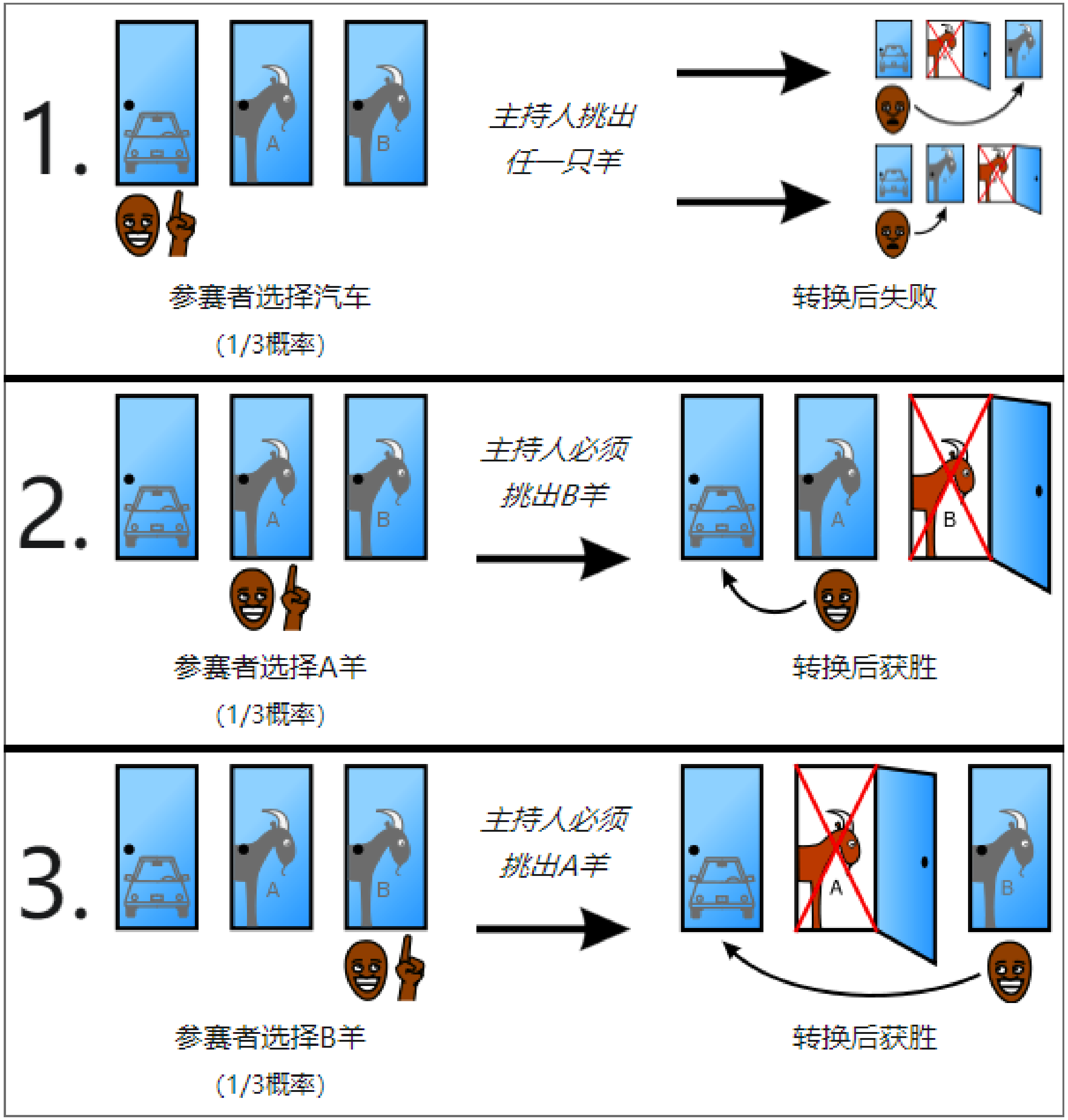
三门问题/ Monty Hall problem/ Парадокс Монти Холла

背景： 出自美国的电视游戏节目Let's Make a Deal

问题： 假设你正在参加一个游戏节目，你被要求在三扇门中选择一扇：其中一扇后面有一辆车；其余两扇后面则是山羊。你选择了一道门，假设是一号门，然后知道门后面有什么的主持人，开启了另一扇后面有山羊的门，假设是三号门。他然后问你：“你想选择二号门吗？”转换你的选择对你来说是一种优势吗？

三门问题/ Monty Hall problem/ Парадокс Монти Холла

解答:



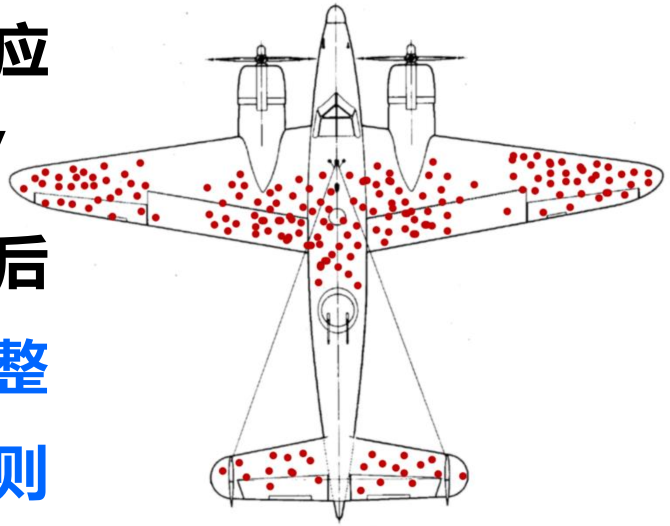
参赛者最初选择时有1/3的相同概率选择汽车、A羊和B羊，转换后的获胜概率为2/3。

幸存者偏差/ survivorship bias/ Систематическая ошибка выжившего

- 是一种逻辑谬误。过度关注“幸存了某些经历”的人事物，忽略那些没有幸存的（可能因为无法观察到），造成错误的结论。

军事案例：

第二次世界大战期间的1941年，美国哥伦比亚大学统计学**亚伯拉罕•沃德**教授接受美国海军的要求，运用他在统计方面的专业知识给出关于‘飞机应该如何加强防护，才能降低被炮火击落的几率’的建议。沃德教授针对盟军的轰炸机遭受攻击后的相关数据，进行分析和研究后发现：**机翼是整个飞机中最容易遭受攻击的位置，但是发动机则是最少被攻击的位置。**



幸存者偏差/ survivorship bias/ Систематическая ошибка выжившего

军事案例（cont.)

- 沃德教授给出的结论是‘我们应该强化发动机的防护’
- 但是美国海军指挥官认为‘应该加强机翼的防护，因为这是最容易被击中的位置’。

沃德教授的依据：

- 统计的样本，仅包含没有因敌火射击而坠毁并安全返航的轰炸机。
- 假设所有中弹的弹著点应该会平均分布在机身各处，而能安全返航的轰炸机机身中弹数量较多的区域，是即使被击中也比较不会导致坠机的部位。
- 机翼被击中很多次的轰炸机，大多数仍然能够安全返航。
- 发动机弹孔较少的原因并非真的不容易中弹，而是一旦中弹，其安全返航并生还的可能性就微乎其微

概率论的起源与发展

起源：

17世纪中叶法国贵族梅勒 $\xrightarrow{\text{赌博问题}}$ 帕斯卡(1623-1662)

→ 费马(1601-1665)

↓

荷兰人惠更斯(1629-1695)：1657年<<论赌博中的计算>>

这一时期称为组合概率阶段

成为数学分支：

瑞士人 雅克比-贝努力(1654-1705) → 大数定理(LLN) $\xrightarrow{\text{成为数学分支}}$ 1713年<<猜度术>>

棣莫佛(1667-1754): <<分析杂论>> → 中心极限定理(CLT)(1901年), 乘法原理， 正态分布等。

蒲丰(1707-1788)：蒲丰问题 → 几何概率

拉普拉斯(1749-1827)：1812《概率分析理论》 → 概率的古典定义

泊松(1781-1840)：推广了大数定理，提出了Poisson分布等.

↓

19世纪后期，极限理论成为概率研究的热点问题，
其中切比雪夫及其学生马尔可夫为代表的圣彼得堡学派做出巨大贡献。

↓

法国人 贝特朗于1899年提出了著名的贝特朗悖论

公理化阶段：

俄国人伯恩斯坦，奥地利人冯-米西斯(1883-1953)对概率论的严格化做了初步的尝试

↓

1905年测度论诞生------波雷尔(1881-1956)，勒贝格(1875-1941)

↓

1933年，柯尔莫哥洛夫(1903-1987)<<概率论基础>> → 现代公理化体系

- **天时地利人和：**
- **莫斯科的 泛函分析 学派**
- **（测度论的发展）**

Андрей Николаевич Колмогоров

当代的发展：随机过程与随机分析---概率主流

莱维(1886-1971): 独立增量过程；　辛钦(1894-1959)：　平稳过程

杜布(1910-2004)，Meyer：鞅论

In 1954, at the age of 39, Kiyoshi Itô gave a series of lectures at the Institute for Advanced Study in Princeton. At the time, the stochastic calculus he had developed was already known in the West.

伊藤清(1915-2008): 1942-1944年定义了对布朗运动的随机积分，

1987年获得沃尔夫奖

Black-Scholes期权定价模型：1973年首次在政治经济学杂志(Journal of Political Economy)发表，　1997年获诺贝尔经济学奖

彭实戈(1947-):

1995年“倒向随机微分方程”获得国家自然科学二等奖(一等奖空

许宝禄(1910-1970)，陈希孺(1934-2005)，严加安(1941---)

马志明(1948----)，陈木法 (1946---)

概率论/ Probability theory/ Теория вероятностей

- **主要数学结论：大数定律 和 中心极限定理**
- **主要概念：随机变量 和随机变量的 概率分布**

基本概念

随机试验（Random Experiment）：对随机现象做出的观察与科学实验。

在概率论中把符合下面三个特点的试验叫做随机试验：

1. 可以在相同的条件下重复的进行。
2. 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果。
3. 进行一次试验之前不能确定哪一个结果会出现。

随机试验通常用E表示。

基本概念

样本空间（Sample space）：随机试验所有的基本可能结果构成的集合称Ω。Ω的元素为**样本点（Sample point）**。

事件（Event）是试验中“人们感兴趣的结果”构成的集合，是Ω的子集。各种不同的事件的总体构成一个事件集合，称为**事件域** F。

事件是随机的。赋予事件一个出现可能性的度量值，称为**概率**（Probability）。

“可能性的度量值”是“宏观”意义下（即大数量的情形下）的比例值，由**相对频率**（Relative frequency）来计算，

$$P(A) \approx \frac{\text{试验中}A\text{出现的次数}}{\text{总试验次数}} = \frac{n_A}{n} \qquad (n\text{很大})$$

概率公理： 任何事件 A 的概率满足：

（1） 非负性：任取事件 A，$P(A) \geq 0$

（2） 归一性：$P(\Omega)=1$

（3） 可加性：若事件 A, B 互斥，即，$A \cap B = \varnothing$，

则，$P(A \cup B)=P(A)+P(B)$

概率空间/Probability space/ Вероятностное пространство:

понятие, введённое А. Н. Колмогоровым в 30-х годах XX века для формализации понятия вероятности.

Вероятностное пространство — это тройка $(\Omega, \mathfrak{A}, \mathbb{P})$ (иногда обрамляемая угловыми скобками: $\langle , \rangle$), где

- Ω — это произвольное непустое множество, элементы которого называются элементарными событиями, исходами или точками;
- $\mathfrak{A}$ — сигма-алгебра подмножеств Ω, называемых (случайными) событиями;
- $\mathbb{P}$ — вероятностная мера или вероятность, то есть сигма-аддитивная конечная мера, такая что $\mathbb{P}(\Omega) = 1$.

概率空间(Ω, F, P)是一个总测度为1的测度空间（即$P(\Omega)$=1）.

第一项Ω是一个非空集合，有时称作“样本空间”。Ω 的集合元素称作“样本输出”，可写作ω。

第二项F是样本空间Ω的幂集的一个非空子集。F的集合元素称为**事件**Σ。事件Σ是样本空间Ω的子集。集合F必须是一个σ-代数:

1. $\Phi \in \mathcal{F}$;
2. 若$A \in \mathcal{F}$，则$\bar{A} \in \mathcal{F}$;
3. 若$A_n \in \mathcal{F}$，$n = 1, 2, \ldots$，则$\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

(Ω, F)合起来称为可测空间。事件就是样本输出的集合，在此集合上可定义其概率。

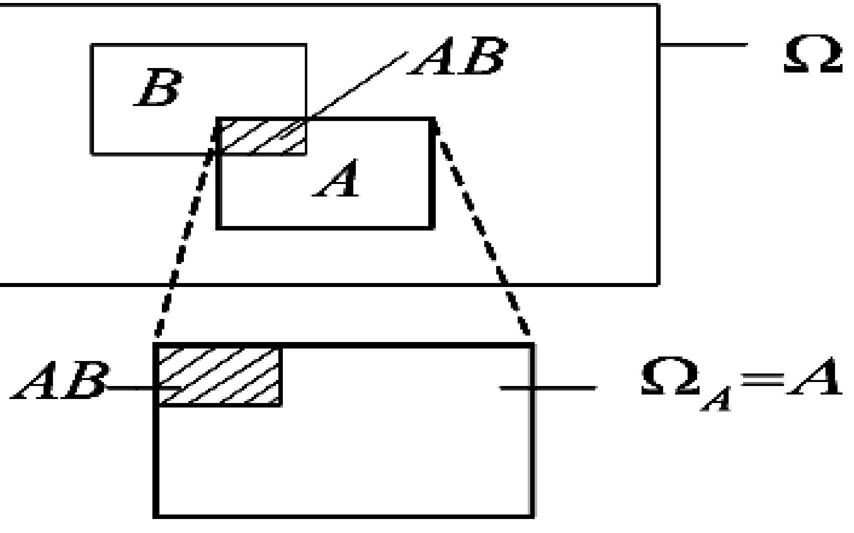
第三项P称为*概率*，或者*概率测度*。这是一个从集合F到实数域R的函数，$P: F \mapsto R$。每个事件都被此函数赋予一个0和1之间的*概率值*。

Условная вероятность

条件事件： $B|A=$ 事件A发生条件下的事件B

条件概率（Conditional probability），

$$P(B|A)=\frac{P(AB)}{P(A)}, \quad P(A)>0$$



Независимость

事件 A 与 B 独立（Independent）等价地定义为

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

多个事件 $A_1, A_2, \cdots, A_n$ 彼此独立，

$$P\left(A_{k_1} A_{k_2} \cdots A_{k_m}\right) = P\left(A_{k_1}\right) P\left(A_{k_2}\right) \cdots P\left(A_{k_m}\right)$$

- **概率论与测度论的区别**

Chain rule/ 链式法则

$$P(A_1A_2\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})$$

全概率公式/ Law of total probability/ Формула полной вероятности

Пусть дано вероятностное пространство $(\Omega, \mathcal{F}, \mathbb{P})$, и полная группа попарно несовместных событий $\{B_i\}_{i=1}^{n} \subset \mathcal{F}$, таких что

1. $\forall i \; \mathbb{P}(B_i) > 0$;
2. $\forall j \neq i \; B_i \cap B_j = \varnothing$;
3. $\bigcup_{i=1}^{n} B_i = \Omega$.

Пусть $A \in \mathcal{F}$ — интересующее нас событие. Тогда получим:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A \mid B_i)\mathbb{P}(B_i).$$

贝叶斯定理/ Bayes' theorem/ Теорема Байеса

Формула Байеса:

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

где

$P(A)$ — априорная вероятность гипотезы A (смысл такой терминологии см. ниже);
$P(A \mid B)$ — вероятность гипотезы A при наступлении события B (апостериорная вероятность);
$P(B \mid A)$ — вероятность наступления события B при истинности гипотезы A;
$P(B)$ — полная вероятность наступления события B.

在贝叶斯定理中，每个名词都有约定俗成的名称：

- $P(A|B)$是已知B发生后，A的条件概率。也由于得自B的取值而被称作A的后验概率。
- $P(A)$是A的先验概率（或边缘概率）。之所以称为"先验"是因为它不考虑任何B方面的因素。
- $P(B|A)$是已知A发生后，B的条件概率。也由于得自A的取值而被称作B的后验概率。
- $P(B)$是B的先验概率。

按这些术语，贝叶斯定理可表述为：

后验概率 = (似然性*先验概率)/标准化常量

也就是说，后验概率与先验概率和相似度的乘积成正比。

另外，比例$P(B|A)/P(B)$也有时被称作标准似然度（standardised likelihood），贝叶斯定理可表述为：

后验概率 = 标准似然度*先验概率

贝叶斯法则与虚假阳性的病例

- 假设在居民区, 一种罕见的疾病传染1‰的人……
- 检验方法:
  - 有人传染上了这种病，其检验结果有99%的可能呈阳性。
  - 也会产生一些虚假的阳性。未传染上的患者有2%检验也呈阳性。
- 而你恰恰检验呈阳性，你传染该病的概率有多大?

贝叶斯法则与虚假阳性的病例

- 我们依据两个事件工作：

A: 患者得这种病

B: 检验结果呈阳性

- 其检验结果的准确性为:
- $P(A)=0.001$ --1‰的人得这种病
- $P(B|A)=0.99$ --真正得这种病的人检验结果呈阳性的概率为99%
- $P(B|A^C)=0.02$ --没得这种病的人检验结果呈假阳性的概率为2%
- 则$P(A|B)$是多少? 即检验结果呈阳性传染上该病的概率是多少?
- $P(A|B)= P(A) * P(B|A) / (P(A)*P(B|A) + (1-P(A))*P(B|A^C))$

  $= 0.001*0.99 / (0.001*0.99+(1- 0.001)*0.02) =0.047$
- 尽管检验精确度很高, 但事实上, 检验呈阳性的人不到5%患有这种病

Байесовская статистика/
Bayesian statistics/贝叶斯统计

- 胰腺癌检测：基于贝叶斯定理：即使100%的胰腺癌症患者都有某症状，而某人有同样的症状，绝对不代表该人有100%的概率得胰腺癌，还需要考虑先验概率，假设胰腺癌的发病率是十万分之一，而全球有同样症状的人有万分之一，则此人得胰腺癌的概率只有十分之一，90%的可能是是假阳性。

- 不良种子检测：基于贝叶斯定理：假设100%的不良种子都表现A性状，而种子表现A性状，并不代表此种子100%是不良种子，还需要考虑先验概率，假设一共有6万颗不良种子，在种子中的比例是十万分之一（假设总共有60亿颗种子），假设所有种子中有1/3表现A性状（即20亿颗种子表现A性状），则此种子为不良种子的概率只有十万分之三。

Случайная величина/ случайная переменная

在样本空间 Ω 上定义一个单值实函数 $X(\xi)$，称为**随机变量**（Random variable，常缩写为r.v.）。并规定：用 $\{X(\xi)\le x\}$ 的概率来描述 $X(\xi)$ 的概率特性，记为

$$F_X(x)=P\{X(\xi)\le x\}$$

称它为 X 的**分布函数**（Distribution function），或称为累积分布函数（Cumulative distribution function）。

- **Фу́нкция распределе́ния**

Пло́тность вероя́тности

X 概率密度函数（Probability density function）

$$f(x)=\frac{d}{dx}F(x)$$

基本性质为：

1. $f(x)\geq 0,\qquad \int_{-\infty}^{+\infty}f(x)dx=1$

2. $P\{X\in A\}=\int_A f(x)dx$

Случайная величина/ случайная переменная

随机变量不同于普通变量表现在两点上：

- 变量可以有多个取值，并且永远不能预知它到底会取哪个值；
- 变量取值是有规律的，这种规律用概率特性来明确表述；

因此，讨论随机变量就必然要联系到它的取值范围与概率特性。

在描述随机变量的概率特性时：

- 分布函数指明直到x处的累积概率；
- 密度函数适用于连续取值部分。
- 离散变量，常采用分布律；

随机变量的数字特征

- □ 数字特征 之于 随机变量 就类似于 标签 之于 QQ好友

- 随机变量的**分布函数**是对随机变量的概率性质**最完整**的刻画
- 能够刻画随机变量某些方面的性质特征的量 称为 随机变量的 数字特征

随机变量的数字特征

- 一维随机变量的数字特征主要是数学期望与方差
- 多维随机变量主要是协方差与相关系数。

数学期望
mathematical expectation/first moment
Математи́ческое ожида́ние

$$\mathrm{E}(X) = \int_\Omega X \, \mathrm{d}P$$

- X是离散的随机变量，输出值为 $x_1, x_2, \ldots$ 和输出值相应的概率 $p_1, p_2, \ldots$ （概率和为1），则

$$\mathrm{E}(X) = \sum_i p_i x_i$$

- X是连续的随机变量，概率密度函f(x), 则

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x) \, \mathrm{d}x$$

方差/二阶矩/ Variance
Диспе́рсия случа́йной величины́

$$\operatorname{Var}(X)=\operatorname{E}\left[(X-\mu)^2\right]$$

- **方差 亦可当作是随机变量与自己本身的 协方差**

$$\operatorname{Var}(X)=\operatorname{Cov}(X,X)$$

- **方差 = 平方的期望减掉期望的平方**

$$\begin{aligned}\operatorname{Var}(X)&=\operatorname{E}\left[(X-\operatorname{E}[X])^2\right]\\&=\operatorname{E}\left[X^2-2X\operatorname{E}[X]+\operatorname{E}[X]^2\right]\\&=\operatorname{E}\left[X^2\right]-2\operatorname{E}[X]\operatorname{E}[X]+\operatorname{E}[X]^2\\&=\operatorname{E}\left[X^2\right]-\operatorname{E}[X]^2\end{aligned}$$

标准差/均方差/ Standard deviation
Среднеквадрати́ческое отклоне́ние

$$\sigma=\sqrt{\operatorname{E}[(X-\mu)^2]}$$

- **方差主要应用在对离散程度的比较方面**

协方差/ Covariance / Ковариáция

期望值分别为$E(X)=\mu$与$E(Y)=\nu$的两个具有有限二阶矩的实数随机变量X与Y之间的协方差定义为:

$$\operatorname{cov}(X,Y)=\mathrm{E}((X-\mu)(Y-\nu))=\mathrm{E}(X\cdot Y)-\mu\nu.$$

- **如果X 与Y 是统计独立的，那么二者之间的协方差就是0**

$$E(X\cdot Y)=E(X)\cdot E(Y)=\mu\nu$$

相关/ Correlation / Корреля́ция

The population correlation coefficient $\rho_{X,Y}$ between two random variables X and Y with expected values μ_X and μ_Y and standard deviations σ_X and σ_Y is defined as

$$\boxed{\rho_{X,Y}=\operatorname{corr}(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sigma_X\sigma_Y}=\frac{\mathrm{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X\sigma_Y}}$$

Pearson's correlation coefficient

$$\rho_{X,Y}=\frac{\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)}{\sqrt{\mathrm{E}(X^2)-\mathrm{E}(X)^2}\cdot\sqrt{\mathrm{E}(Y^2)-\mathrm{E}(Y)^2}}$$

独立 => 0

离散型r.v.的分布：伯努利分布/0-1分布/ Bernoulli distribution/ Распределéние Бернýлли

- **若试验成功，则随机变量取值为1。**
- **若试验失败，则随机变量取值为0。**
- **记其成功概率为 p，失败概率为 q=1-p。**

其概率质量函数为：

- $$f_X(x) = p^x(1-p)^{1-x} = \begin{cases} p & \text{if } x = 1, \\ q & \text{if } x = 0. \end{cases}$$

$$\mathrm{E}(X) = p$$

$$\mathrm{Var}[X] = pq = p(1-p)$$

离散型r.v.的分布：二项分布/ Binomial distribution / Биномиáльное распределéние

- 是n个独立的 是/非 试验 中成功的次数的离散概率分布，其中每次试验的成功概率为p。
- 当n = 1时，二项分布就是伯努利分布

一般地，如果随机变量X服从参数为n和p的二项分布，我们记$X \sim b(n,p)$或$X \sim B(n,p)$。n次试验中正好得到k次成功的概率由概率质量函数给出：

$$f(k,n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

对于k = 0, 1, 2, ..., n，其中$\binom{n}{k} = \frac{n!}{k!(n-k)!}$

是二项式系数（这就是二项分布的名称的由来），又记为$C(n, k)$，${}_nC_k$，或${}^nC_k$。

累积分布函数可以表示为：

$$F(x;n,p) = \Pr(X \le x) = \sum_{i=0}^{\lfloor x \rfloor} \binom{n}{i} p^i (1-p)^{n-i}$$

其中$\lfloor x \rfloor$是小于或等于x的最大整数。

X的期望值为

$$\mathrm{E}[X] = np$$

方差为

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

离散型r.v.的分布：泊松分布/ Poisson distribution / Распределéние Пуассóна

$X \backsim P(\lambda)$

- **描述单位时间内随机事件发生的次数的概率分布**
- **如某一服务设施在一定时间内受到的服务请求的次数，电话交换机接到呼叫的次数、汽车站台的候客人数、机器出现的故障数、自然灾害发生的次数、DNA序列的变异数等等**
- **二项分布可以看作泊松分布在离散时间上的对应物**

泊松分布的概率质量函数为：

$$P(X=k)=\frac{e^{-\lambda}\lambda^k}{k!}$$

$$E(X)=V(X)=\lambda$$

连续型r.v.的分布：均匀分布/ uniform distribution / непреры́вное равноме́рное распределе́ние

$X \backsim U[a,b]$

概率密度函数:

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

累积分布函数:

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

期望值和中值：是指连续型均匀分布函数的期望值和中值等于区间[a,b]上的中间点。

$$E[X] = \frac{a+b}{2}$$

方差:

$$VAR[X] = \frac{(b-a)^2}{12}$$

连续型r.v.的分布：指数分布/ Exponential distribution/ Экспоненциальное распределение

X ∽ Exp(λ)

- 表示独立随机事件发生的时间间隔
- 比如旅客进入机场的时间间隔、打进客服中心电话的时间间隔、维基百科新条目出现的时间间隔等等。

The probability density function (pdf) of an exponential distribution is

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

The cumulative distribution function is given by

$$F(x;\lambda) = \begin{cases} 1 - e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

$$\mathrm{E}[X] = \frac{1}{\lambda} \qquad \mathrm{Var}[X] = \frac{1}{\lambda^2}$$

连续型r.v.的分布：正态分布/高斯分布/normal distribution / Gaussian distribution/ Норма́льное распределе́ние

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

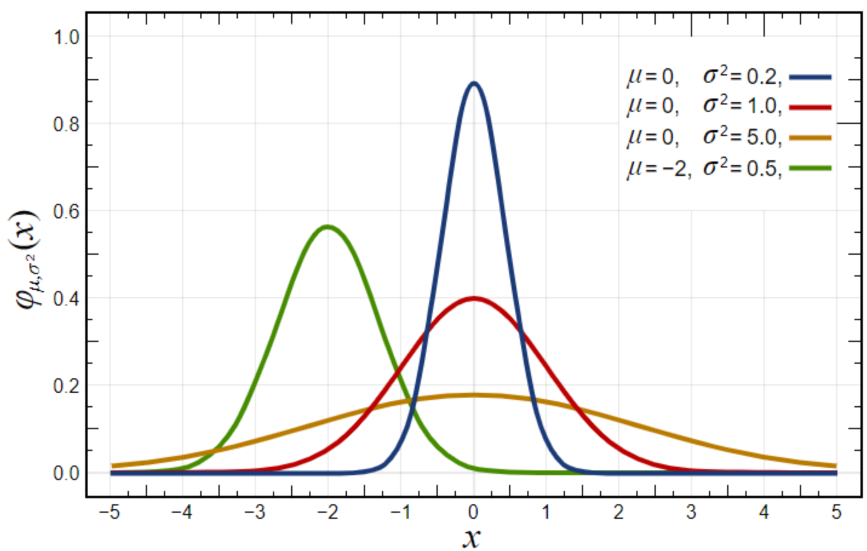
正态分布的概率密度函数均值为μ 方差为σ^2 (或标准差σ)是高斯函数的一个实例：

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

(*请看指数函数以及π.*)

如果一个随机变量X服从这个分布，我们写作 $X \sim N(\mu, \sigma^2)$. 如果$\mu = 0$并且$\sigma = 1$，这个分布被称为**标准正态分布**，这个分布能够简化为

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



MatLab

分布名称	产生随机数	均值与方差
二项分布	binornd	binostat
泊松分布	poissrnd	poissstat
离散均匀分布	unidrnd	unidstat
均匀分布	unifrnd	unifstat
指数分布	exprnd	expstat
正态分布	normrnd	normstat
瑞利分布	raylrnd	raylstat
χ^2 方分布	chi2rnd	chi2stat

MatLab

分布名称	密度函数值	分布函数值
二项分布	binopdf	binocdf
泊松分布	poisspdf	poisscdf
离散均匀分布	unidpdf	unidcdf
均匀分布	unifpdf	unifcdf
指数分布	exppdf	expcdf
正态分布	normpdf	normcdf
瑞利分布	raylpdf	raylcdf
χ^2 方分布	chi2pdf	chi2cdf

实验 1：随机数的产生与测量：产生 $\lambda=1.0$ 的 10000 个指数分布随机数，测量均值、方差与概率密度。

解：

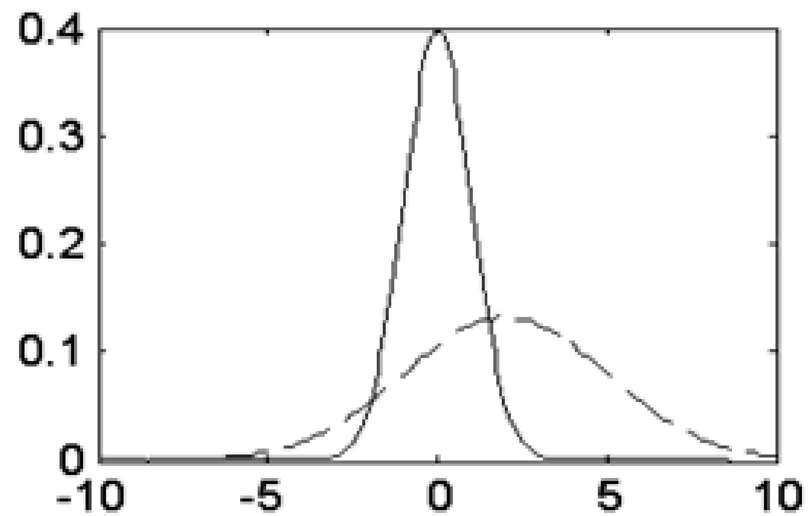
```
Xi_array=exprnd(0.5, 1,10000);
mean(Xi_array)                ; %   ans =2.0019
var(Xi_array)                 ; %   ans =4. 0939
hist(Xi_array)
```

实验 2：绘制 $N(0,1)$ 与 $N(2,3)$ 的密度函数曲线。

解：

```
x=-10:0.1:10;
curve1=normpdf(x,0,1
);
plot(x, curve1,'k')

curve2=normpdf(x,2,3
);
plot(x, curve2,'k-')
```



大数定律/ Law of large numbers / Закон больших чисел

- **大数定律说: 如果统计数据足够大，那么事物出现的频率就能无限接近他的期望值。**

- **Опр.: сходимость по вероятности 依概率收敛**

数学表达式就是，对于 $\forall \varepsilon > 0$ 有:

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

(即越界是小概率事件)

或 $\lim\limits_{n\to\infty} P\{|X_n - X| \leq \varepsilon\} = 1$

(即绝大部分是在界内的)

切比雪夫大数定律/ Chebyshev / Чебышёв

2.1前提条件

① $X_1, X_2, ..., X_n$ 相互独立（**注意：不要求同分布！**）

② DX_i 存在且一致有上界（严谨表述： $\exists C$ 使 $DX_i \leq C$ 对一切 $i \geq 1$ 成立)

2.2结论

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} EX_i$$

2.3数学意义

算数平均值依概率收敛于数学期望

伯努利大数定律/ Bernoulli's Theorem / Борелевский закон больших чисел

我们再来看Bernoulli大数定律，这是概率论历史上第一个极限定理，属于Chebyshev的一种特殊情况，可以由Chebyshev推出。

3.1前提条件

① μ_n 是n重Bernoulli实验中事件A的发生次数

②每次试验A发生的概率为p

所谓**n重Bernoulli实验**，就是"不成功便成仁"，独立重复地进行n次实验，成功了就是1，不成功就是0。所以我们把这两个前提条件照着Chebyshev翻译一下就是：

$X_1, X_2, \ldots, X_n$ **相互独立且都服从于参数为p的0-1分布**

$$X_i \sim \begin{pmatrix} 1 & 0 \\ p & 1-p \end{pmatrix}$$

3.2结论

$$\frac{1}{n}\mu_n \xrightarrow{P} p$$

3.3数学意义

频率依概率收敛于统计概率

辛钦大数定律 (弱大数定律)/ Khinchin's law (weak law of large numbers)/ Слабый закон

4.1前提条件

① $X_1, X_2, \ldots, X_n$ 相互独立且同分布

② $EX_i = \mu$ 存在

4.2结论

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

4.3数学意义

算数平均值稳定于数学期望的确切解释

强大数定律/ strong law of large numbers/ Усиленный закон (Теорема Колмогорова)

强大数定律指出，样本均值以概率1收敛于期望值。

$$\overline{X}_n \xrightarrow{\text{a.s.}} \mu \quad \text{as} \quad n \to \infty$$

即

$$P\left(\lim_{n\to\infty} \overline{X}_n = \mu\right) = 1$$

中心极限定理/

Central limit theorem /

Центральная предельная теорема

- **中心极限定理说明:**

在适当的条件下，大量相互独立随机变量的均值经适当标准化后依分布收敛于正态分布

林德伯格-莱维中心极限定理/ Lindeberg–Lévy CLT/ ЦПТ Линдеберга

5.1前提条件

① $X_1, X_2, ..., X_n$ 相互独立且同分布

② $EX_i = \mu,\ DX_i = \sigma^2$ 存在

5.2结论

则对于 $\forall x \in R$ 有:

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

也即 X_i 求和近似服从正态分布:

$$\sum_{i=1}^{n} X_i \sim N\left(n\mu, n\sigma^2\right)$$

当n足够大的时候，我们可以把任何一个奇奇怪怪的分布，搞成一个正态分布!

棣莫佛－拉普拉斯定理/ De Moivre–Laplace theorem/ Локальная теорема Муавра — Лапласа

6.1前提条件

$$Y_n \sim B(n,p)\,,\ 0<p<1,\ n\geq 1$$

别看这个前提条件就一句话，但数学语言就是这样，蕴含着丰富的信息：**De Moivre-Laplace定理其实就是Levi-Lindeberg的特殊情况。**

你看 Y_n 服从二项分布，那么它不就是① $Y_1, Y_2, \ldots, Y_n$ 相互独立且同分布，且② $EY_i=\mu,\ DY_i=\sigma^2$ 存在的吗，也就是说，完全满足了Levi-Lindeberg的两个条件——实际上，我们就是在**把一个二项分布，尝试转为正态分布去研究。**

6.2结论

那么自然而言地，我们可以套用Levi-Lindeberg的结论：

对于 $\forall x\in R$ 有：

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

也即：

$$Y_n \sim N(np, np(1-p))$$