《统计与数据分析》
Statistics & Data Analysis

§2 概率理论基础
(Part III)

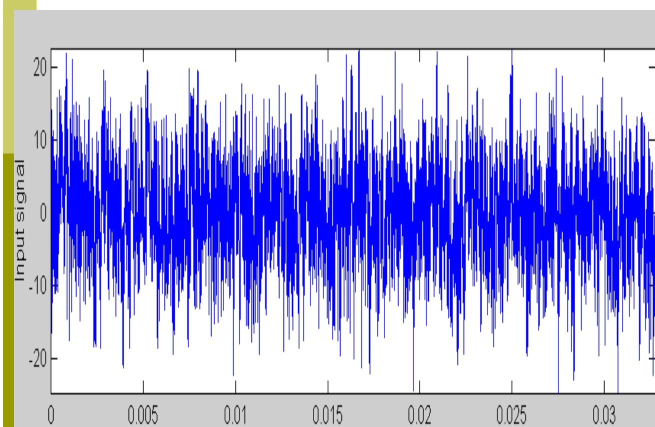
Zhu Huaiqiu
@Peking University

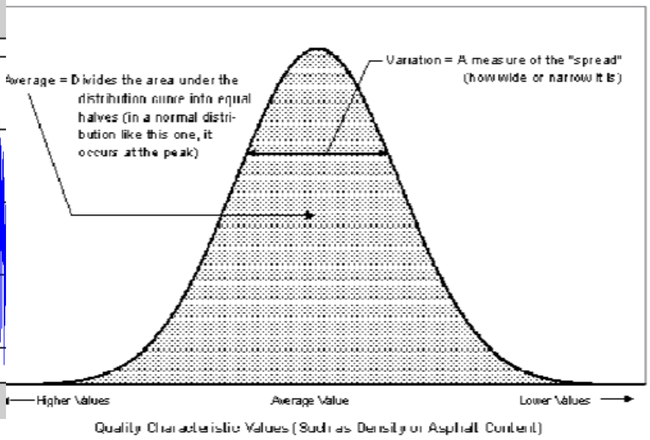
§2.3 随机变量的期望特征与极限性质

How can we describe the properties of a random variable?

Distribution

Expectation properties





§2.3.1 随机变量的期望值

一、离散型随机变量的期望值

定义

设离散型随机变量X的分布律为 $P(X=x_i)=p_i$，**i=1, 2, ...,**

若级数 $\sum_{i=1}^{\infty} x_i p_i$ **绝对收敛，则称**

$$\sum_{i=1}^{\infty} x_i p_i \qquad \text{（加权平均值）}$$

为X的期望值(expected value)或期望(expectation)或均值(mean)，记作

$$E(X)=\sum_i x_i p_i$$

常用离散型分布的期望值：

（1）两点分布：

随机变量X~B(1, p)，E(X)=p

（2）二项分布：

随机变量X~B(n, p)，E(X)=np

（3）几何分布：

随机变量X~G(p)，E(X)=1/p

（4）Poisson分布：

随机变量X~P(λ)，E(X)=λ

【Example 2.36】St. Petersburg paradox

18世纪30年代，Bernoulli提出一个问题，掷硬币赌博，假定为公平的赌博，某赌徒的方案为：开始赌注为1元，若输，下一次赌注加倍，如此连续直至第一次赢，且只赢回该次的赌注。这种方案能保证他最后赢得1元。

请分析这一方案的可行性。

二、连续型随机变量的期望值

定义

设X为密度函数为f(x)的连续型随机变量，若积分

$$\int_{-\infty}^{+\infty} xf(x)dx$$

绝对收敛，称该积分为X的**期望值**(expected value)或**期望**(expectation)或**均值**(mean)，记作

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx \quad \text{（加权平均值）}$$

常用连续型分布的期望值：

（1）均匀分布：

随机变量X~U[a, b]，其密度函数为

$$f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b]\\ 0, & otherwise\end{cases},$$

其中a<b为常数，则：

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\frac{a+b}{2}$$

（2）指数分布：

随机变量X~E(λ)，密度函数为 $f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & otherwise\end{cases}$,

则： $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\frac{1}{\lambda}$

（3）Gamma分布：

随机变量X~*Γ*(α, λ)，密度函数为 $f(x)=\begin{cases}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\lambda x}, & x>0,\\ 0, & x\leq 0,\end{cases}$

则： $E(X)=\int_{0}^{\infty}\frac{\lambda^{\alpha}}{\Gamma(\alpha)}x^{\alpha}e^{-\lambda x}dx=\frac{\alpha}{\lambda}$

（4）正态分布:

随机变量X~N(μ, σ²), 密度函数为 $f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},-\infty<x<+\infty$

则: $E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\mu$

三、随机变量函数的期望值

定理

设X是随机变量，Y=g(X)是X的函数。

（1）X为离散型随机变量，其分布律为 $P(X=x_i)=p_i$，i=1, 2, ...，当级数 $\sum_{i=1}^{\infty}g(x_i)p_i$ 绝对收敛时，有:

$$E(Y)=E\left(g(X)\right)=\sum_{i}g(x_i)p_i$$

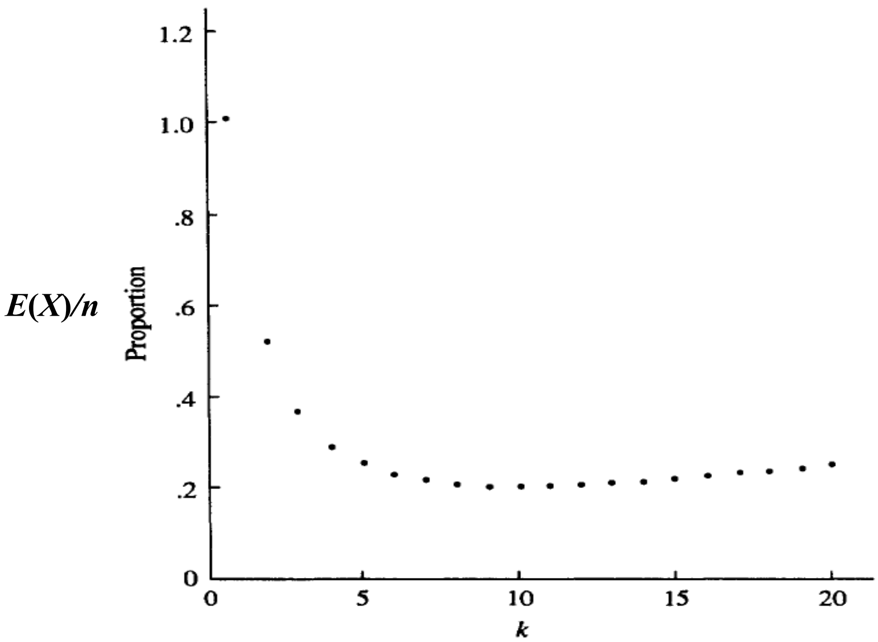
（2）X为连续型随机变量，其密度函数为f(x)，若积分 $\int_{-\infty}^{+\infty}g(x)f(x)dx$ 绝对收敛时，有:

$$E(Y)=E\left(g(X)\right)=\int_{-\infty}^{+\infty}g(x)f(x)dx$$

【Example 2.37】设某产品的市场需求量是随机变量X（单位：吨），服从[2000, 4000]上的均匀分布。设该产品每销售1吨可获利3万元；但若销售不出囤积仓库，1吨需保管费1万元。问应该生产多少这种产品，使得获利最大？

The expected value is a calculation that serves as the best prediction of a value. In financial terms, it is the probability-weighted average value of all possible outcomes. Understanding the expected value of a possible future event allows us to make mathematically sound decisions.

【Example 2.38】化验n份血液采样样本，来检测某一罕见疾病（发病率p=1%），假定化验方法有足够高的灵敏性。可采用三种方案：（1）逐份化验；（2）将每份样本分成两份：先将每个半份混在一起做1次合并化验（pooling test），若呈阴性，表明都没有，化验完毕；若呈阳性，再将另外的n个半份逐份化验；（3）推广方法：假定将n份分为m组，每组有k份，即n=mk，对每一组采用第2种办法，即先混合做1次合并化验若呈阳性，进一步进行逐份化验。试讨论这些方法所需化验次数的期望值。



四、期望值的性质

<u>性质1</u>
若C为常数，则E(C)=C

<u>性质2</u>
若C为常数，X为随机变量，则E(CX)=CE(X)

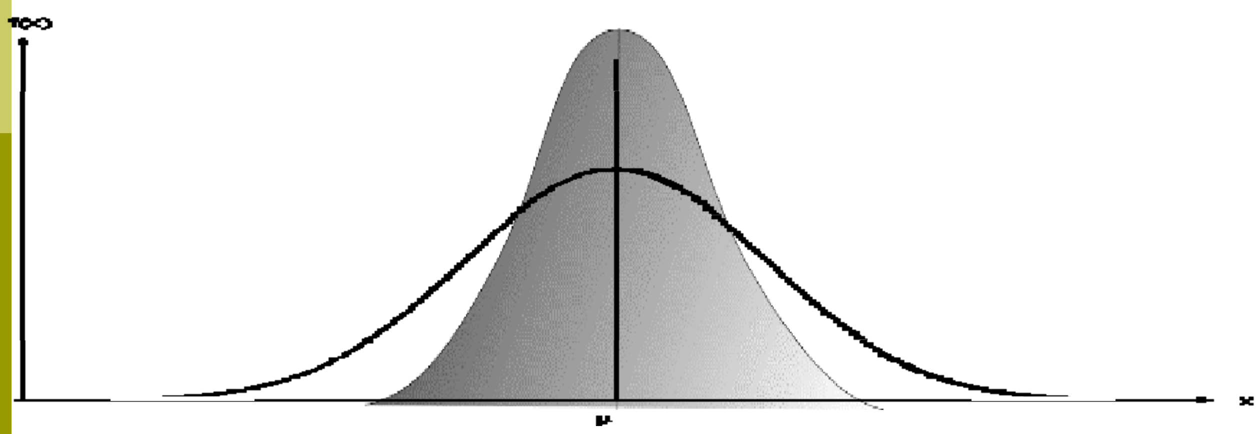
<u>性质3</u>
若X，Y为任意两个随机变量，则有：E(X+Y)=E(X)+E(Y)。（可推广到n个随机变量）

<u>性质4</u>
若X, Y相互独立，则有E(XY)=E(X)E(Y)。
（可推广到n个变量）

§2.3.2　方差和标准差

<u>位置参数</u>（location parameter）

<u>标度参数</u>（scale parameter）



一、定义

定义

若对于随机变量X， $E\left\{\left[X-E(X)\right]^2\right\}$ 存在，

则称之为X的**方差**（variance），记为D(X)，即

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

并称 $\sigma(x)=\sqrt{D(X)}$ 为X的**标准差**（standard deviation）。

计算方差的重要公式：

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

二、常见分布的方差

（1）两点分布：

随机变量X~B(1, p)，E(X)=p，D(X)=p(1－p)

（2）二项分布：

随机变量X~B(n, p)，E(X)=np，D(X)=np(1－p)，n≥2

（3）Poisson分布：

随机变量X~P(λ)，E(X)=λ，D(X)= λ

（4）指数分布：

随机变量X~E(λ)

则：$E(X)=\frac{1}{\lambda}$ $D(X)=\frac{1}{\lambda^2}$

（5）正态分布：

随机变量X~N(μ, σ²)

则：$E(X)=\mu,$ $D(X)=\sigma^2$

三、方差性质的讨论

性质1

若C为常数，则D(C)=0。

性质2

若X为随机变量，C为常数，则D(CX)=C²D(X)。

性质3

若X_1, X_2是任意两个随机变量，则有

$$D(X_1\pm X_2)=D(X_1)+D(X_2)\pm 2E\{[X_1-E(X_1)][X_2-E(X_2)]\}$$
$$=D(X_1)+D(X_2)\pm 2[E(X_1X_2)-E(X_1)E(X_2)]$$

性质4

若X_1, X_2两个随机变量相互独立，则有 $D(X_1\pm X_2)=D(X_1)+D(X_2)$

四、方差和测量误差

测量误差的统计模型

设某物理量的真实值为x_0，测量值X可表示为：

$$X = x_0 + \beta + \varepsilon$$

其中，β为系统误差，ε为随机误差。且有$E(\varepsilon)=0$，$D(\varepsilon)=\sigma^2$

$$E(X) = x_0 + \beta, \qquad D(X) = \sigma^2$$

通常用均方误差MSE（mean squarred error）来表示测量误差：

$$MSE = E\left[(X - x_0)^2\right] = \beta^2 + \sigma^2$$

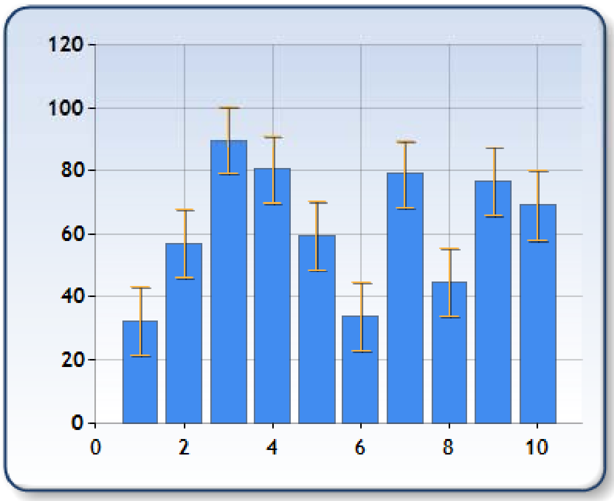
系统误差（systematical error）：
系统误差的大小在测量过程中是不变的，可以用计算或实验方法求得，即是可预测的，并且可修正或调整使其减少。

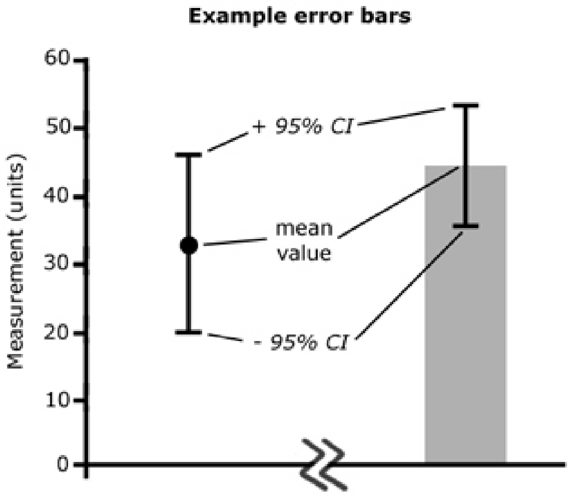
偶然误差（random error）：
在相同条件下，对同一物理量进行多次测量，由于各种偶然因素，会出现测量值时而偏大、时而偏小的误差现象，这种类型的误差叫做偶然误差。在确定的测量条件下，对同一物理量进行多次测量，并且用它的算术平均值作为该物理量的测量结果，能够比较好地减少偶然误差。

偶然误差的统计规律：
（1）绝对值相等的正的与负的误差出现机会相同；
（2）绝对值小的误差比绝对值大的误差出现的机会多；
（3）误差不会超出一定的范围。

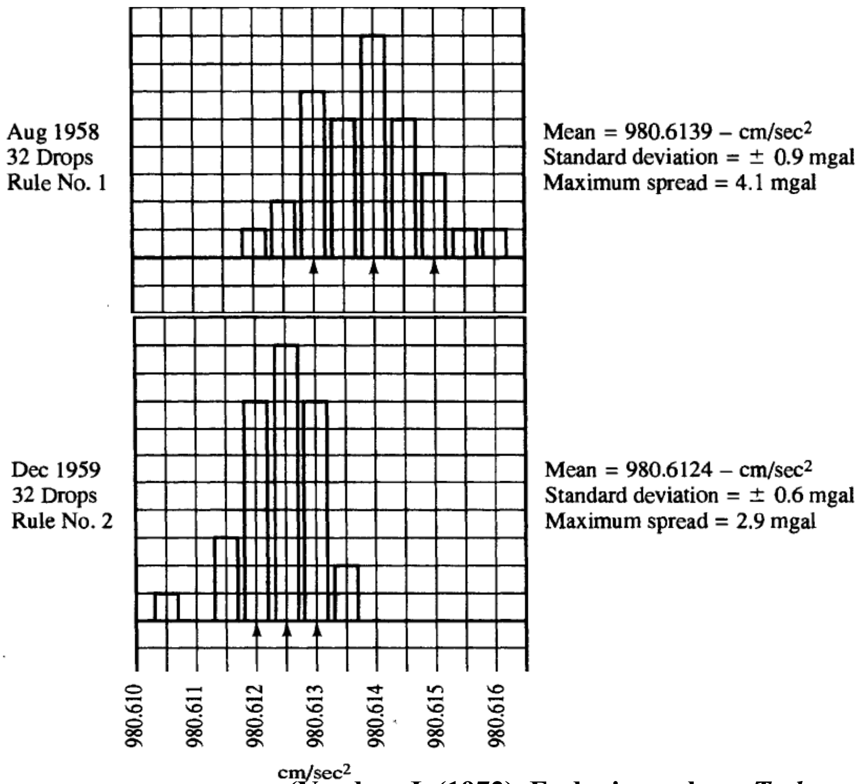
Error bar：表示与随机误差σ相关的量

Often represent one standard deviation of uncertainty, one standard error, or a certain confidence interval (e.g., a 95% interval). So the measure selected should be stated explicitly in the graph or supporting text.





【Example 2.39】重力加速度g的测量：Preston-Thomas等1958-1959年在加拿大渥太华同一地点采用两种不同方法分别测量32次。



(Youden, J. (1972). Enduring values. *Technometrics*, 14: 1-11)

§2.3.3 协方差和相关系数

一、协方差

定义

设(X, Y)为二维随机变量，若 $E\{[X-E(X)][Y-E(Y)]\}$ 存在，则称其为X与Y的协方差（covariance），记为cov(X, Y)或σ_{xy}，即

$$\sigma_{xy}=\operatorname{cov}(X,Y)=E\{[X-E(X)][Y-E(Y)]\}$$

当cov(X, Y)=0时，称X与Y不线性相关，简称不相关。

关于协方差的讨论

（1）(X, Y)为二维离散型随机变量时，设其联合分布律为：

$$p_{ij}=P(X=x_i,Y=y_j),\quad i,j=1,2,\ldots$$

则有：

$$\operatorname{cov}(X,Y)=\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}\{[x_i-E(X)][y_i-E(Y)]p_{ij}\}$$

（2）(X, Y)为二维连续型随机变量时，设其联合密度函数为f(x, y)，则有：

$$\operatorname{cov}(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[x-E(X)][y-E(Y)]f(x,y)dxdy$$

（3）常用计算协方差的公式:

$$\mathrm{cov}(X,Y)=E(XY)-E(X)E(Y)$$

(4) $\mathrm{cov}(X,X)=D(X)$

（5）X, Y的不相关、X, Y相互独立之间的联系：相互独立，必然不相关；反之不成立!

【Example 2.40】

二维连续型随机变量**(X, Y)**的联合密度函数为

$$f(x,y)=\begin{cases}\frac{1}{4}(1-x^3y+xy^3), & |x|<1,|y|<1\\ 0, & otherwise\end{cases}$$

考察**X, Y**是否相关？是否相互独立？

协方差的性质

性质1　$cov(X, Y)=cov(Y, X)$

性质2　$cov(aX, Y)=a\,cov(X, Y)$
$cov(X, aY)=a\,cov(X, Y)$
a, b是常数

性质3　$cov(X+Y, Z)=cov(X, Z)+cov(Y, Z)$

性质4　$cov(aX\pm b, Y)=a\,cov(X, Y)$，a, b是常数

二、相关系数

定义

若二维随机变量(X, Y)的协方差存在，且有D(X)>0, D(Y)>0，则称

$$\frac{\operatorname{cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

为X与Y的相关系数(correlation coefficient)，记为$\rho(X, Y)$或ρ_{xy}，即

$$\rho_{xy}=\rho(X,Y)=E\left[\frac{X-E(X)}{\sqrt{D(X)}}\cdot\frac{Y-E(Y)}{\sqrt{D(Y)}}\right]=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

若$\rho(X, Y)=0$，则称X与Y不相关。

定理

对若二维随机变量(X, Y)，有：

(1) $\rho(X,Y)=\rho(Y,X)$

(2) $|\rho(X,Y)|\le 1$

(3) $|\rho(X,Y)|=1$ **的充要条件是：存在不为零的常数k, b，使得** $P(Y=kX+b)=1$

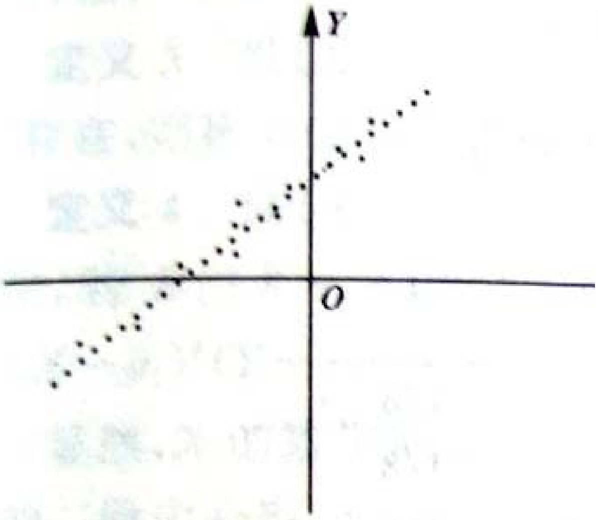
相关系数的讨论：

（1）“X与Y相互独立”是“X与Y不相关”的充分条件，而非必要条件；

（2）当(X, Y)服从二维正态分布时，“X与Y相互独立”与“X与Y不相关”等价，互为充分必要条件；

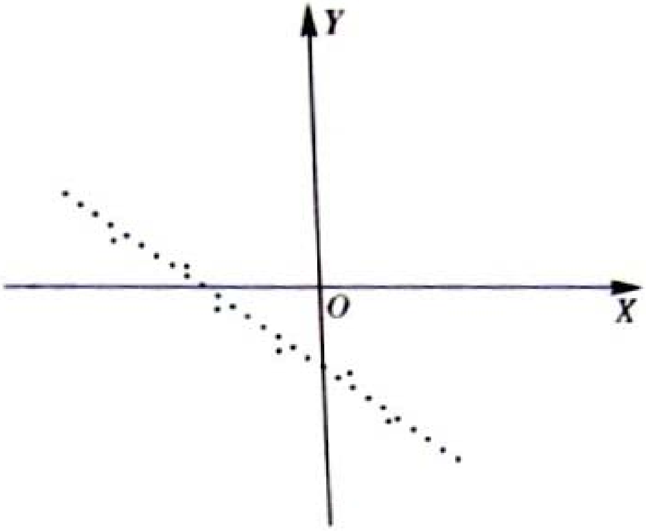
（3）相关系数ρ(X, Y)是一个无量纲量，不依赖于(X, Y)的量纲，因此比cov(X, Y)更为常用；

（4）相关系数ρ(X, Y)反映二维随机向量(X, Y)的两个分量X与Y间的线性联系，但不能反映X与Y之间的非线性联系。



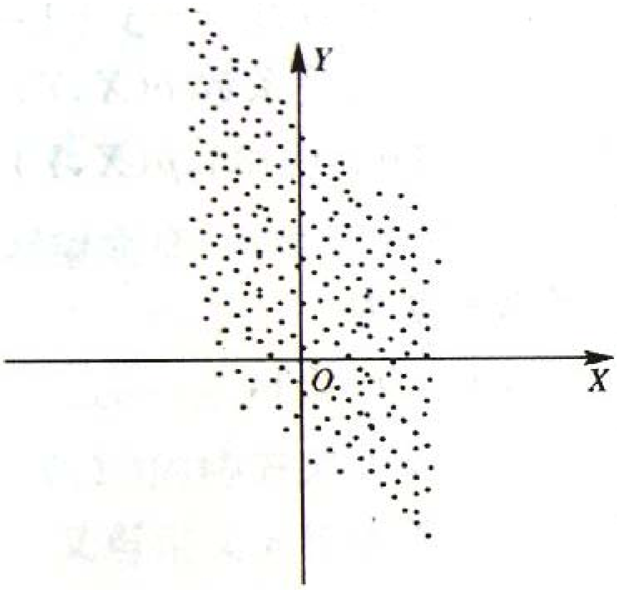
(a) $\rho(X,Y)\approx 1$

X 与 Y 线性联系密切



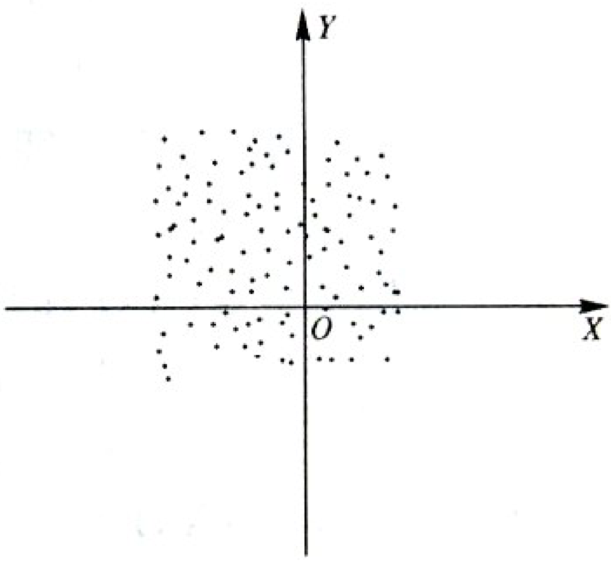
(b) $\rho(X,Y)\approx -1$

X 与 Y 的线性联系密切



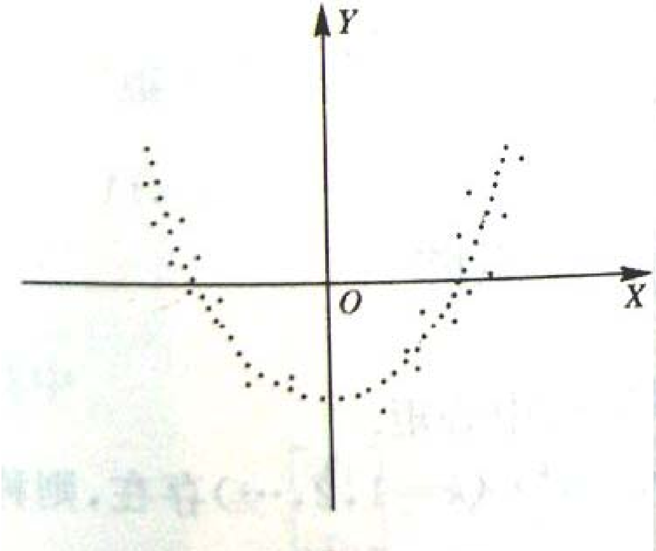
(c) $-1<\rho(X,Y)<0$

X 与 Y 线性联系不密切

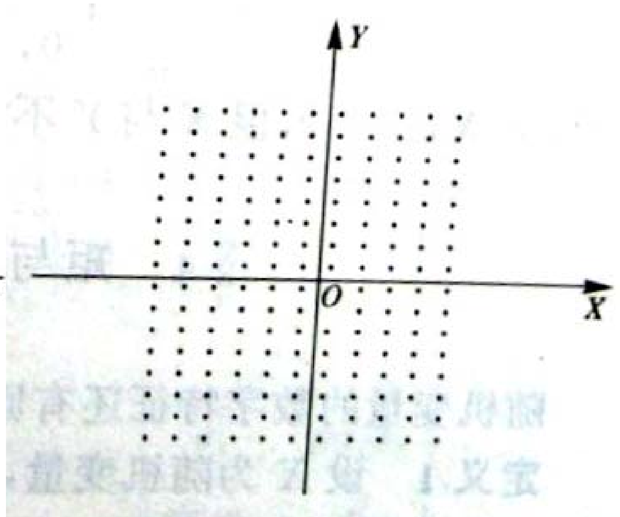


(d) $\rho(X,Y)\approx 0$

X 与 Y 无线性联系



(e) $\rho(X,Y)=0$
X与Y无线性联系



(f) $\rho(X,Y)=0$
X与Y独立，X与Y无线性联系

三、协方差矩阵

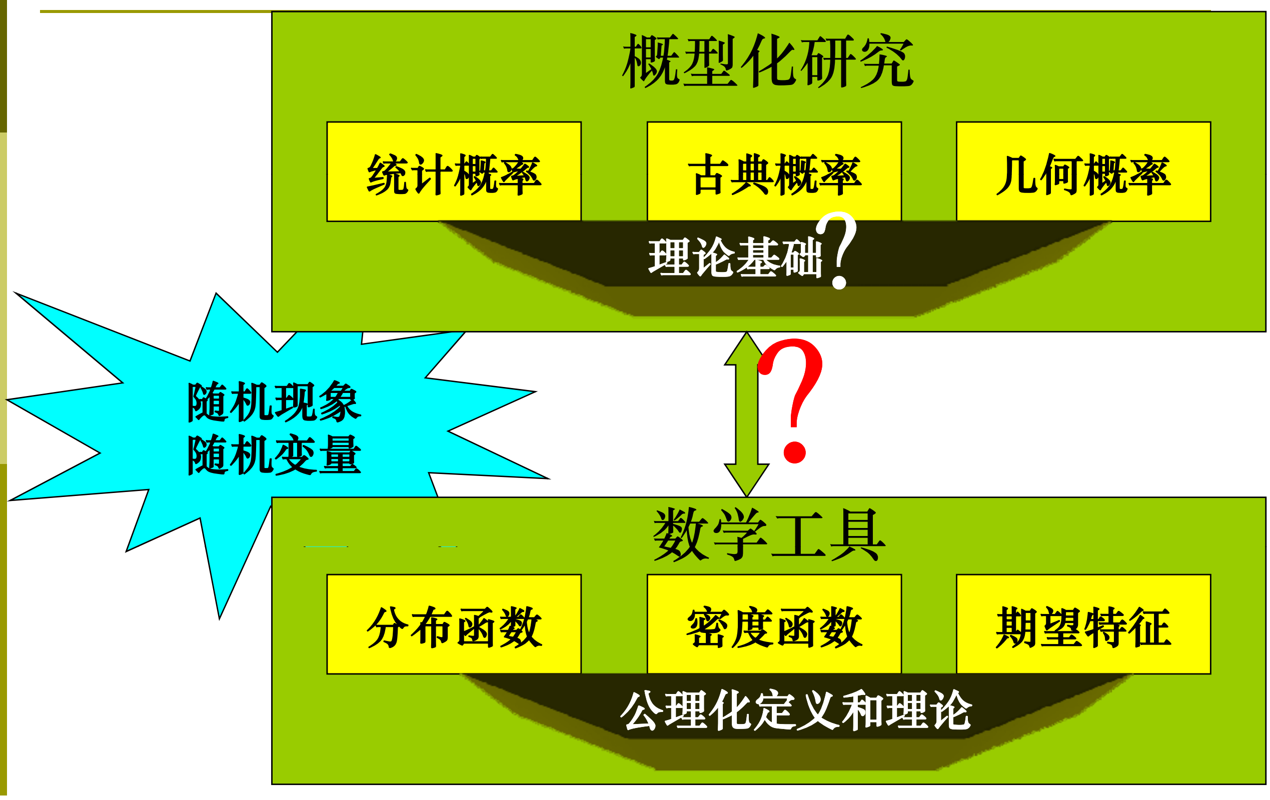
定义 对二维随机变量(X, Y)，称(E(X), E(Y))为(X, Y)的**期望向量**或**均值向量**；称矩阵

$$\begin{bmatrix} D(X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & D(Y) \end{bmatrix} = \begin{bmatrix} \mathrm{cov}(X,X) & \mathrm{cov}(X,Y) \\ \mathrm{cov}(Y,X) & \mathrm{cov}(Y,Y) \end{bmatrix}$$

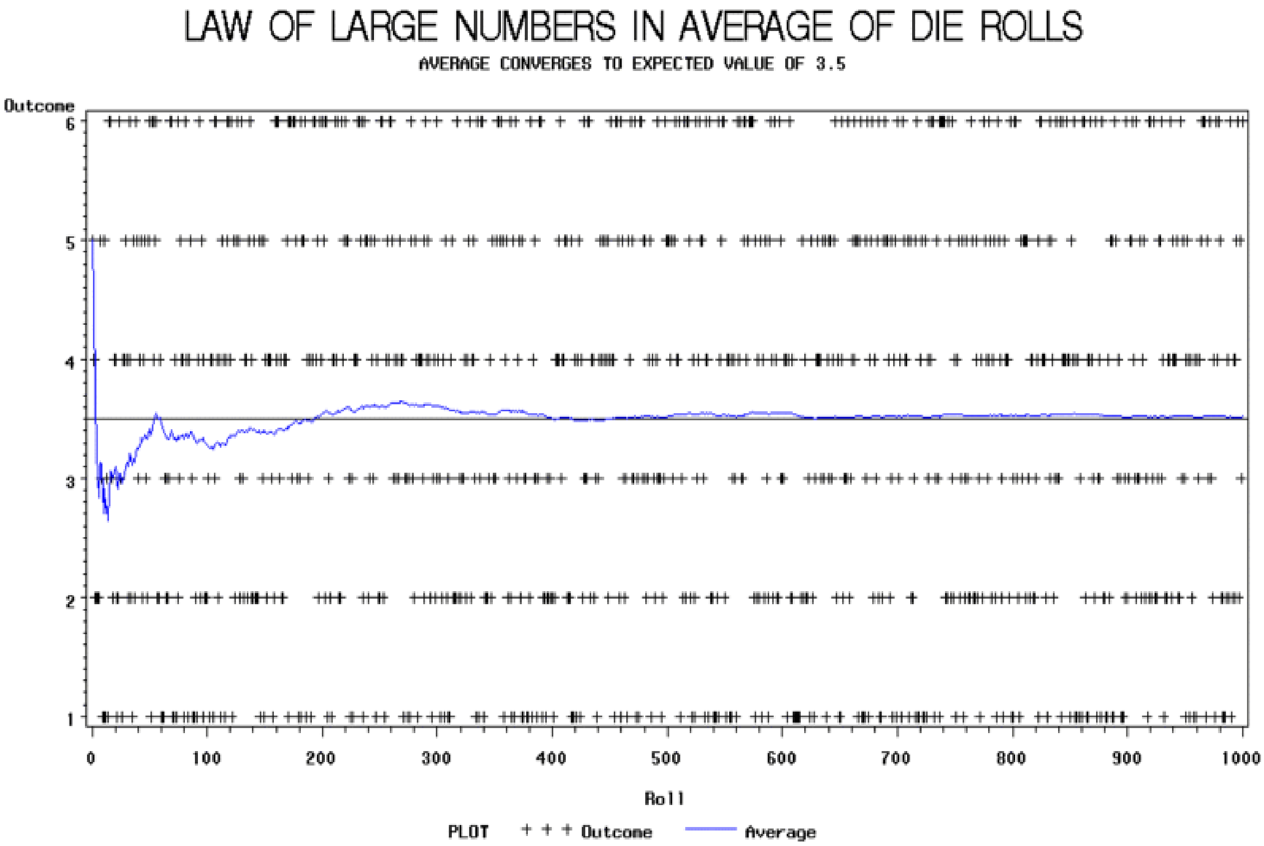
为(X, Y)的**协方差矩阵**（covariance matrix）。

由此可定义n维随机向量的$(X_1, X_2, ..., X_n)$的协方差矩阵：

$$\begin{bmatrix} D(X_1) & \mathrm{cov}(X_1,X_2) & \dots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & D(X_2) & \dots & \mathrm{cov}(X_2,X_n) \\ \dots & \dots & \dots & \dots \\ \mathrm{cov}(X_n,X_1) & \mathrm{cov}(X_n,X_2) & \dots & D(X_n) \end{bmatrix} = \begin{bmatrix} \mathrm{cov}(X_1,X_1) & \mathrm{cov}(X_1,X_2) & \dots & \mathrm{cov}(X_1,X_n) \\ \mathrm{cov}(X_2,X_1) & \mathrm{cov}(X_2,X_2) & \dots & \mathrm{cov}(X_2,X_n) \\ \dots & \dots & \dots & \dots \\ \mathrm{cov}(X_n,X_1) & \mathrm{cov}(X_n,X_2) & \dots & \mathrm{cov}(X_n,X_n) \end{bmatrix}$$



§ 2.3.4 大数定律（Law of large number）



定理 大数定律（LLN, Law of large number）：

设$X_1, X_2, ..., X_i, ...$为一独立的随机变量序列，且$E(X_i)=\mu$，$D(X_i)=\sigma^2$。令

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

则对于任意的$\varepsilon>0$，都有：

$$\lim_{n\to\infty}\left\{P\left(\left|\overline{X}_n-\mu\right|>\varepsilon\right)\right\}=0$$

此为Chebyshev大数定律。

定理 大数定律（LLN, Law of large number）：

（Chebyshev大数定律的特殊情形）

设μ_n是n次独立试验中事件A出现的次数，而p是事件A在每次试验中出现的概率，则对任意的$\varepsilon>0$，均有

$$\lim_{n\to\infty}\left\{P\left(\left|\frac{\mu_n}{n}-p\right|>\varepsilon\right)\right\}=0.$$

此为Bornoulli大数定律。

讨论：

（1）The LLN "guarantees" stable long-term results for random events;

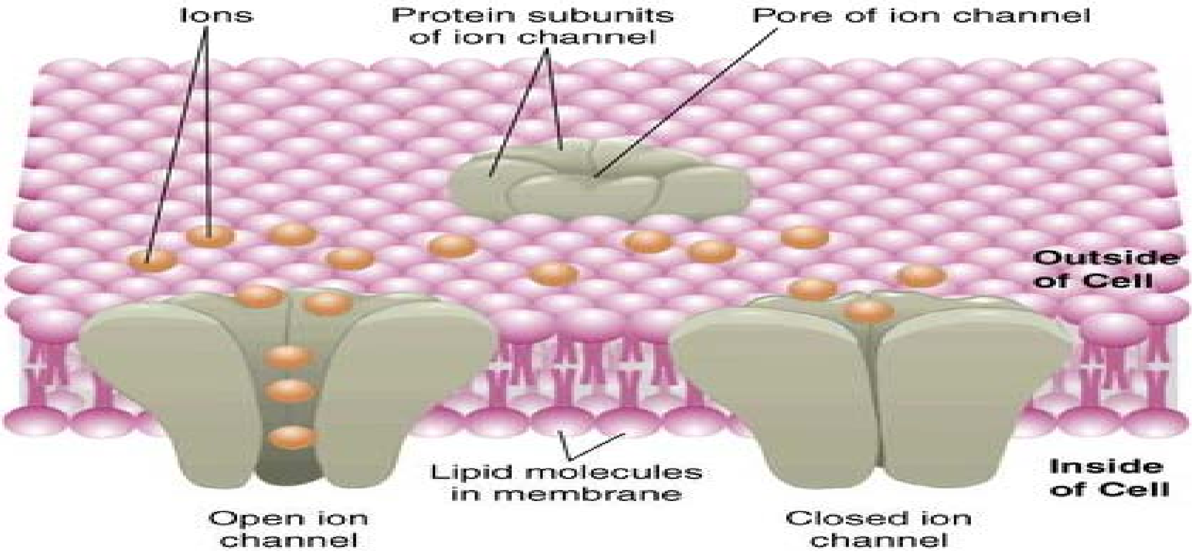
（2）The LLN only applies (as the name indicates) when a ***large number*** of observations are considered;

（3）对于随机变量序列$\{Z_n\}$，若存在实数α，使得

$$\lim_{n\to\infty}\left\{P\left(\left|Z_n-\alpha\right|>\varepsilon\right)\right\}=0$$

我们说$\{Z_n\}$依概率收敛到α

【Example 2.41】细胞膜的离子通道电流量的测定:
肌肉和神经细胞的膜上有大量的离子通道供所需的离子穿过。一般认为在平衡状态下只有极少数的通道是打开的，每一通道的打开和闭合状态是随机且相互独立的。设共有m个通道，某一通道某一时刻处在打开状态的概率为p（很小），电流量为c。已知实验可测得任一时刻的总电流量S，要求推导单个通道的电流量c值。



§ 2.3.5 中心极限定理（Central limit theorem）

定理 **独立随机变量的中心极限定理**

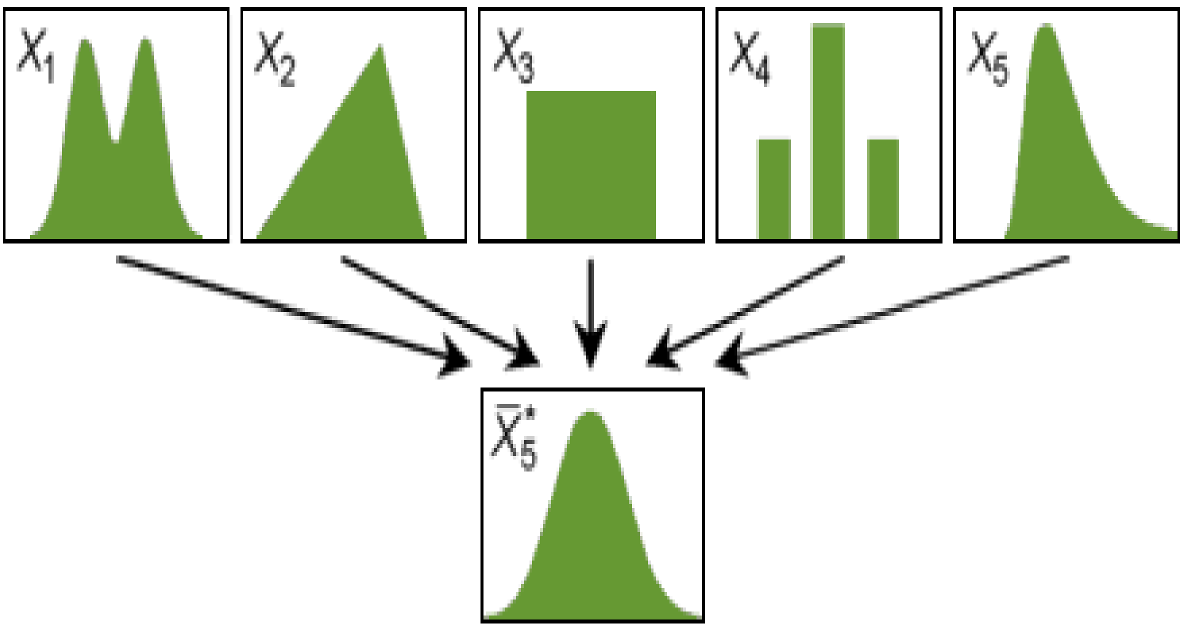
设$\{X_k\}$为相互独立的随机变量序列，且X_k均有有限的期望值和方差：

$$E(X_k)=\mu_k,\qquad D(X_k)=\sigma_k^2,\qquad k=1,2,\ldots$$

若对于任意的实数x，有：

$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k-\sum_{k=1}^{n}\mu_k}{\sqrt{\sum_{k=1}^{n}\sigma_k^2}}\le x\right)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

All five independent distributions have mean 0 and standard deviation 1 and are dramatically non-normal. They were selected arbitrarily, but their normalized average is approximately normal.



定理 独立同分布随机变量的中心极限定理

设$\{X_k\}$为相互独立且服从相同分布的随机变量序列，且X_k均有有限的期望值和方差：

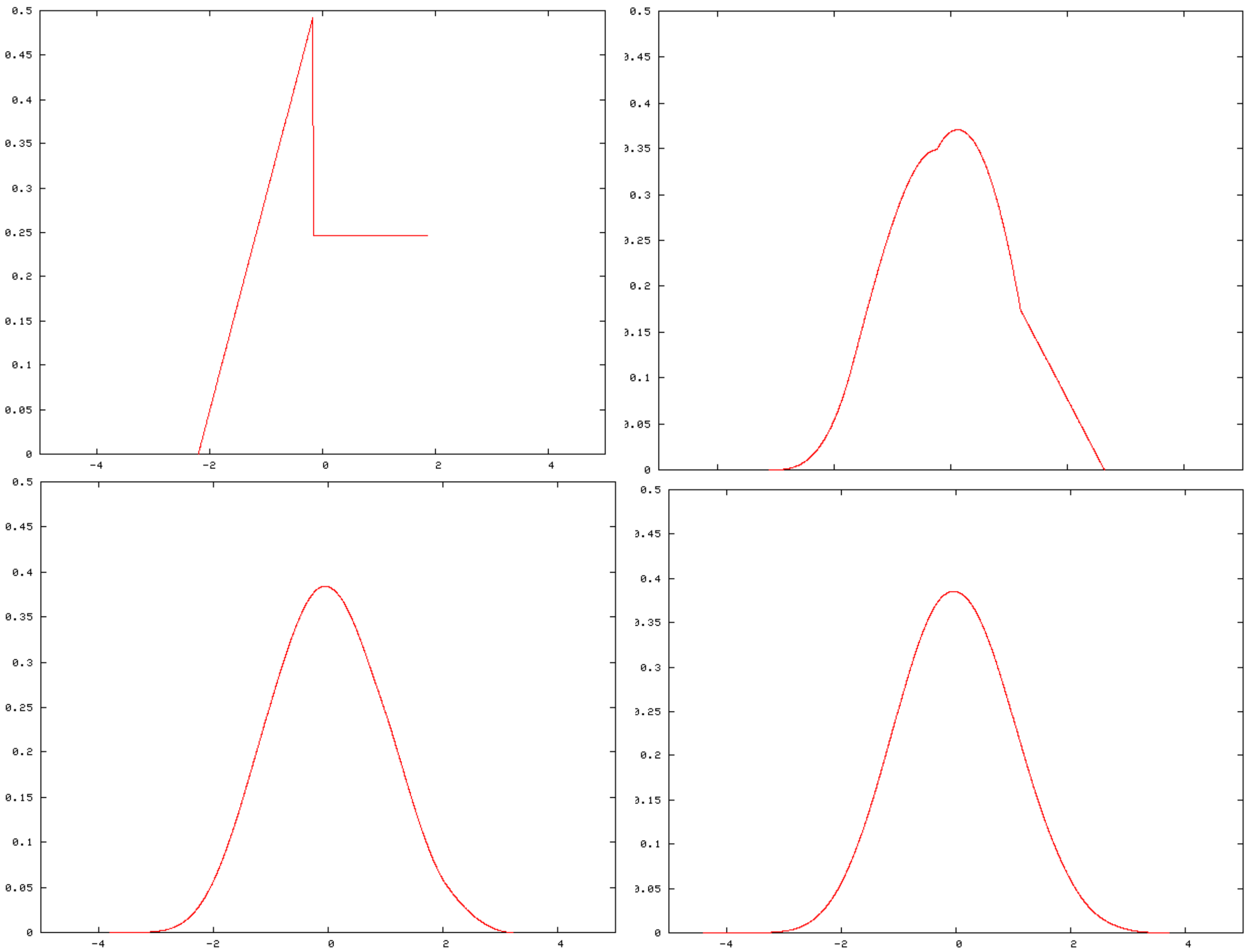
$$E(X_k)=\mu, \qquad D(X_k)=\sigma^2, \qquad k=1,2,\dots$$

若对于任意的实数x，有：

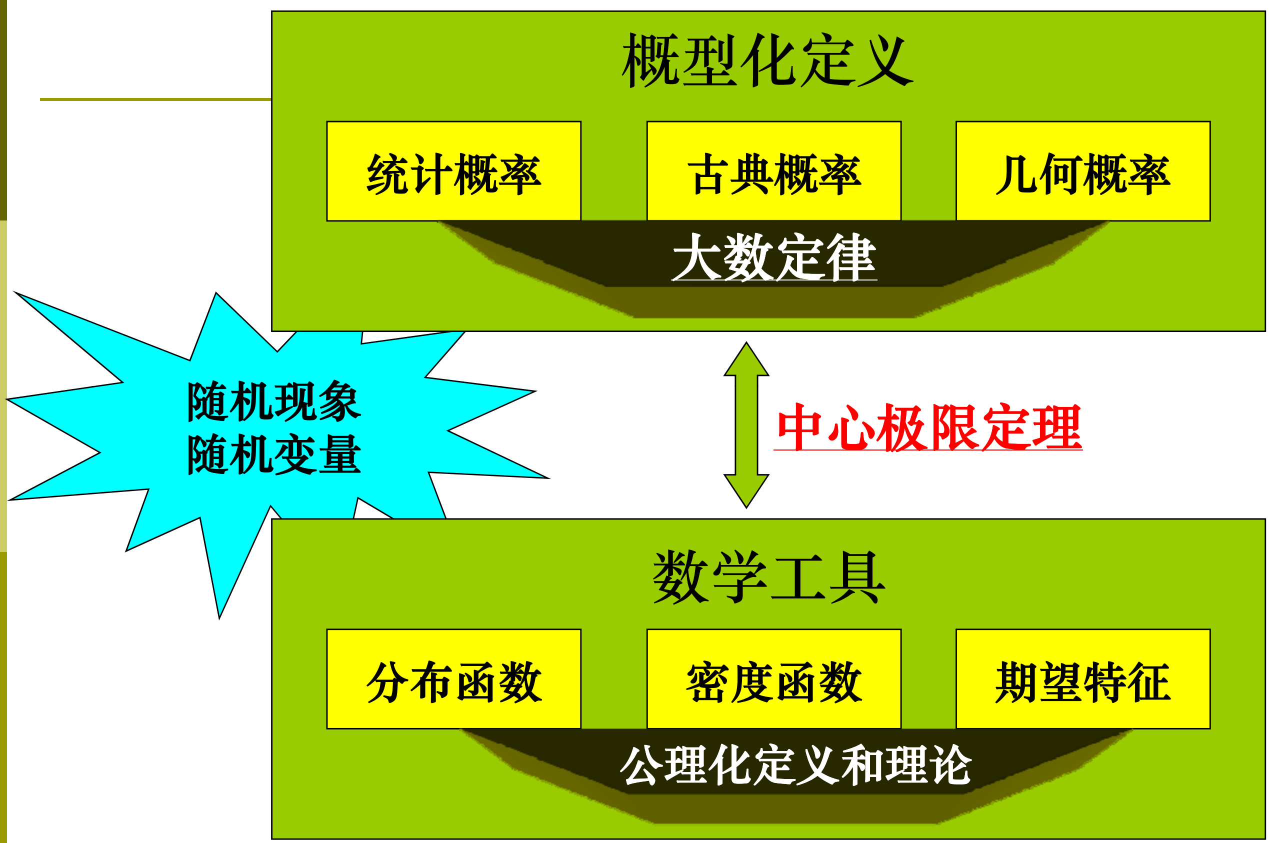
$$\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k - n\mu}{\sigma\sqrt{n}} \le x\right) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt$$

【Example 2.42】A probability density function is shown in the first figure below. Then the densities of the sums of two, three, and four independent identically distributed variables, each having the original density, are shown in the following figures.

If the original density is a piecewise polynomial, as it is in the example, then so are the sum densities, of increasingly higher degree. Although the original density is far from normal, the density of the sum of just a few variables with that density is much smoother and has some of the qualitative features of the normal density.



Overview:



概型化定义
统计概率
古典概率
几何概率
大数定律
随机现象
随机变量
中心极限定理
数学工具
分布函数
密度函数
期望特征
公理化定义和理论