第二节　　中心极限定理

- 中心极限定理
- 例题
- 课堂练习
- 小结　布置作业

中心极限定理的客观背景

在实际问题中许多随机变量是由相互独立随机因素的综合（或和）影响所形成的.

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素（如瞄准、空气阻力、炮弹或炮身结构等）综合影响的.每个随机因素的对弹着点（随机变量和）所起的作用都是很小的.那么弹着点服从怎样分布哪 ？

自从高斯指出测量误差服从正态分布之后，人们发现，正态分布在自然界中极为常见.

高斯

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成，而每一个别因素对这种综合影响中所起的作用不大. 则这种随机变量一般都服从或近似服从正态分布.

现在我们就来研究独立随机变量之和所特有的规律性问题.

当n无限增大时，这个和的极限分布是什么呢？

由于无穷个随机变量之和可能趋于∞，故我们不研究n个随机变量之和本身而考虑它的标准化的随机变量. 即考虑随机变量$X_k\ (k=1,\cdots n)$的和$\sum_{k=1}^{n} X_k$的标准化的随机变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}}$$

讨论Y_n的极限分布是否为标准正态分布

在概率论中，习惯于把和的分布收敛于正态分布这一类定理都叫做中心极限定理.

一、中心极限定理

列维—林德伯格中心极限定理

【定理1】（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n, \cdots$相互独立，服从同一分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2$ $(k=1,2,\cdots)$，则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$$Y_n = \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$对于任意x满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \le x \right\}$$

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

注　1、定理表明，独立同分布的随机变量之和 $\sum_{k=1}^{n} X_k$，当n充分大时，随机变量之和与其标准化变量分别有

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2)\ ; \qquad \frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1).$$

2、独立同分布中心极限定理的另一种形式可写为

$$\overline{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n) \quad \text{或} \quad \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1)$$

其中 $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$

3、虽然在一般情况下，我们很难求出 $\sum_{k=1}^{n} X_k$ 的分布的确切形式，但当n很大时，可以求出近似分布.

【定理2】（李雅普诺夫(Liapounov)定理)

设随机变量$X_1, X_2, \cdots, X_n \cdots$相互独立，它们具有数学期望和方差：

$$E(X_k) = \mu_k,\quad D(X_k) = \sigma_k^{\ 2}, (k = 1,2,\cdots)$$

记 $$B_n^{\ 2} = \sum_{k=1}^{n} \sigma_k^{\ 2}$$

若存在正数δ，使得当$n \to \infty$时，

$$\frac{1}{B_n^{\ 2+\delta}} \sum_{k=1}^{n} E\left\{\left|X_k - \mu_k\right|^{2+\delta}\right\} \to 0$$

则随机变量之和$\sum\limits_{k=1}^{n} X_k$的标准化变量：

$$Z_n = \frac{\sum_{k=1}^{n} X_k - E(\sum_{k=1}^{n} X_k)}{\sqrt{D(\sum_{k=1}^{n} X_k)}} = \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n}$$

的分布函数$F_n(x)$对于任意x，满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{ \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} \mu_k}{B_n} \le x \right\}$$

$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

$$= \Phi(x)$$

请注意：

1、定理中随机变量之和$\sum_{k=1}^{n} X_k$及其标准化变量Z_n在n很大时,分别近似服从

$$\sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N\left(\sum_{k=1}^{n}\mu_k, B_n^2\right); \quad Z_n \overset{\text{近似地}}{\sim} N(0,1)$$

2、随机变量X_k无论服从什么分布，只要满足定理条件，随机变量之和$\sum_{k=1}^{n} X_k$，当n很大时，就近似服从正态分布，这就是为什么正态分布在概率论中所占的重要地位的一个基本原因.

中心极限定理发展史

- 第一版是法国数学家棣莫弗发现的，他在1733年发表的论文中使用正态分布去估计大量抛掷硬币出现正面次数的分布。
- 这个超越时代的成果险些被历史遗忘，所幸著名法国数学家拉普拉斯在1812年发表的巨著Théorie Analytique des Probabilités中拯救了这个默默无名的理论。拉普拉斯扩展了棣莫弗的理论，指出二项分布可用正态分布逼近。但同棣莫弗一样，拉普拉斯的发现在当时并未引起很大反响。
- 直到十九世纪末中心极限定理的重要性才被世人所知。
- 1901年，俄国数学家李雅普诺夫用更普通的随机变量定义中心极限定理并在数学上进行了精确的证明。

【定理3】(棣莫佛—拉普拉斯(De Movire Laplace)定理)

设随机变量η_n(n=1, 2, ···)服从参数$n, p(0<p<1)$的二项分布，则对任意x，有

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

【证】由第四章知识知可将η_n分解成为n个相互独立、服从同一(0−1)分布的诸随机变量$X_1, X_2, \cdots X_n$之和，即有

$$\eta_n = \sum_{k=1}^{n} X_k$$

其中$X_k (k = 1,2,\cdots,n)$的分布律为

$$P\{X_k = i\} = p^i (1-p)^{1-i}, i = 0,1$$

由于$E(X_k)=p, D(X_k)=p(1-p) \quad k=1,2,\cdots,n)$,

由定理1得

$$\lim_{n\to\infty} P\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\} = \lim_{n\to\infty} P\{\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \le x\}$$

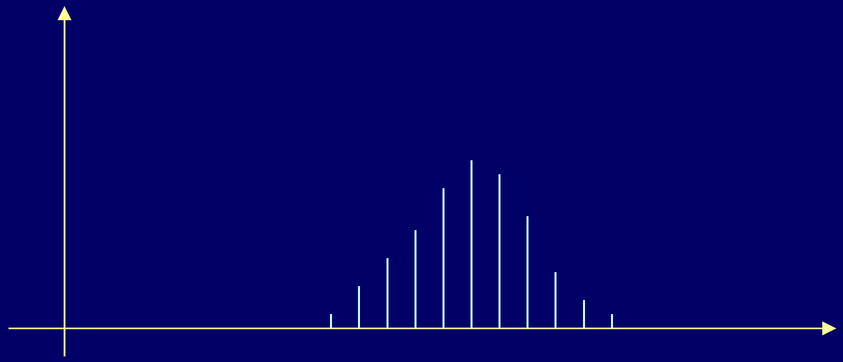
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

定理表明：当n很大，$0<p<1$是一个定值时（或者说，$np(1-p)$也不太小时），二项变量η_n的分布近似正态分布 $N(np, np(1-p))$.

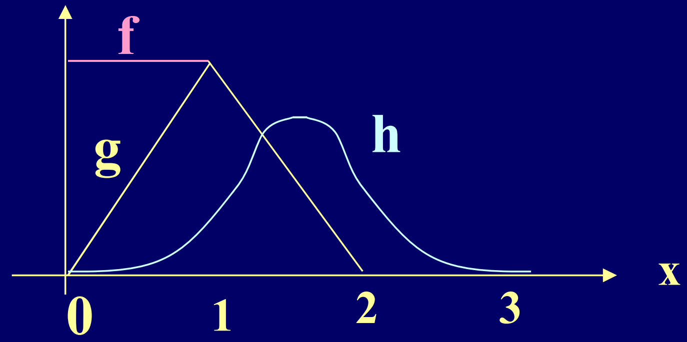
即

$$\eta_n \overset{\text{近似地}}{\sim} N(np, np(1-p))$$

下面演示不难看到中心极限定理的客观背景



例:20个0-1分布的和的分布



几个(0,1)上均匀分布的和的分布

$X_1 \sim f(x)$　　$X_1+X_2 \sim g(x)$

$X_1+X_2+X_3 \sim h(x)$

■ 几个(0,1)上均匀分布R.V. X_i的和的分布

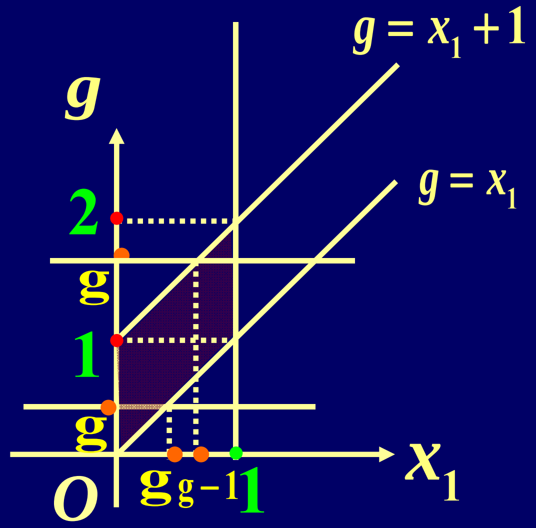
$$X_i \sim U(0,1)$$

$$f(x_i)=\begin{cases}1, & 0<x<1\\ 0, & \text{其它}\end{cases}$$

$$G=X_1+X_2$$

$$f_g(g)=\int_{-\infty}^{\infty} f_{X_1}(x_1)f_{X_2}(g-x_1)dx_1$$

$$=\begin{cases}\int_0^{g} dx_1=g\,, & 0\le g<1,\\ \int_{g-1}^{1} dx_1=2-g\,, & 1\le g<2,\\ 0\,, & \text{其它}.\end{cases}$$



■ 几个(0,1)上均匀分布R.V. X_i的和的分布

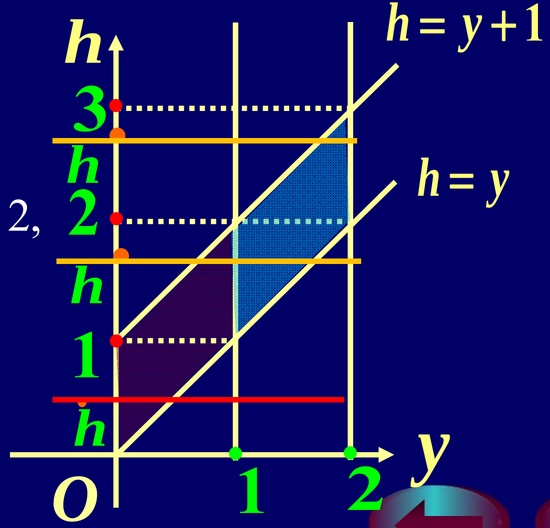
$$X_i \sim U(0,1) \qquad G = X_1 + X_2$$

$$f(x_i) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{其它} \end{cases}$$

$$f_g(g) = \begin{cases} \int_0^{g} dx_1 = g, & 0 \le g < 1, \\ \int_{g-1}^{1} dx_1 = 2 - g, & 1 \le g < 2, \\ 0, & \text{其它.} \end{cases}$$

$$H = X_1 + X_2 + X_3, \text{令} Y = X_1 + X_2$$

$$f_h(h) = \begin{cases} \int_0^{h} y\,dy = \dfrac{h^2}{2}, & 0 \le h < 1, \\ \int_{h-1}^{1} y\,dy + \int_1^{h} (2-y)\,dy = -h^2 + 3h - \dfrac{3}{2}, & 1 \le h < 2, \\ \int_{h-1}^{2} (2-y)\,dy = \dfrac{h^2}{2} - 3h + \dfrac{9}{2}, & 1 \le h < 2, \\ 0, & \text{其它.} \end{cases}$$

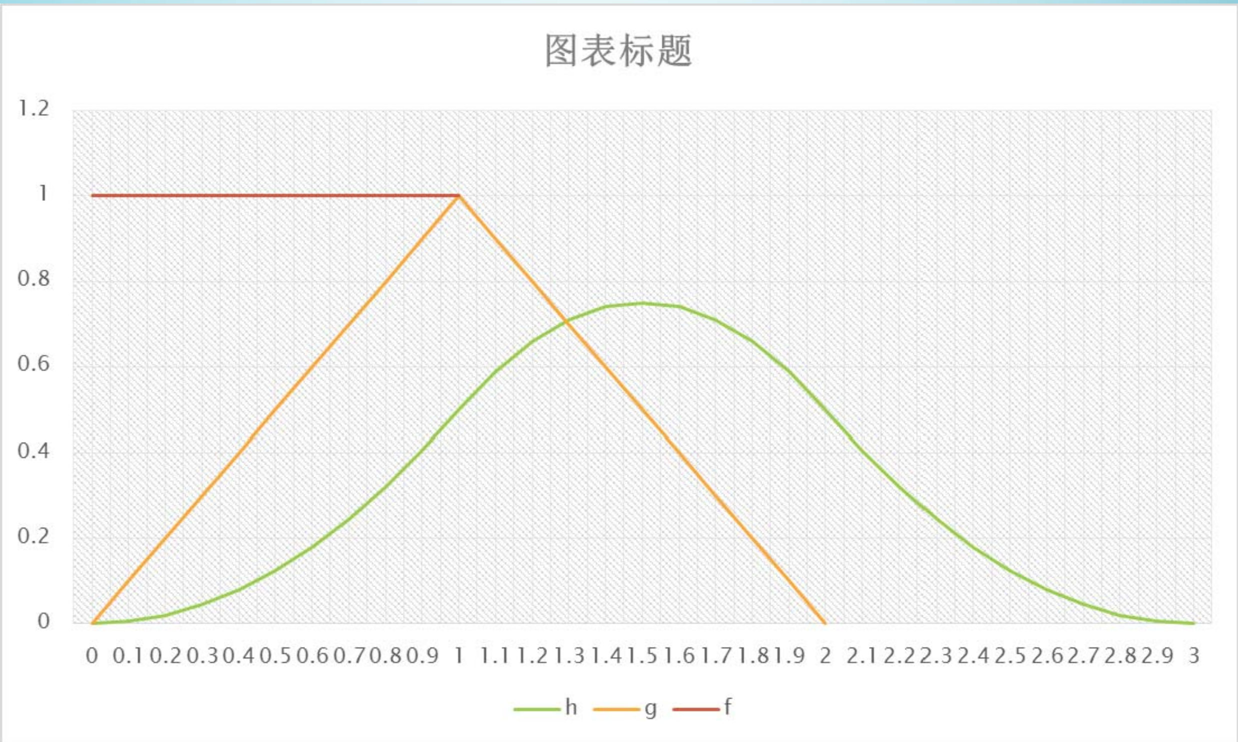


- 几个(0,1)上均匀分布R.V. X_i的和的分布

$$F = X_i \sim U(0,1)$$

$$G = X_1 + X_2$$

$$H = X_1 + X_2 + X_3, \text{令} Y = X_1 + X_2$$



二、例题

【例1】根据以往经验，某种电器元件的寿命服从均值为100小时的指数分布. 现随机地取16只，设它们的寿命是相互独立的. 求这16只元件的寿命的总和大于1920小时的概率.

解：　设第i只元件的寿命为X_i, i=1,2, …,16

$$X_k \sim E(1/\theta), 则E(X_k)=\theta, D(V_k)=\theta^2$$

由题给条件知，诸X_i独立，且$E(X_i)$=100，$D(X_i)$=10000

16只元件的寿命的总和为 $Y=\sum_{k=1}^{16} X_k$

依题意，所求为$P(Y>1920)$　$E(Y)$=1600，$D(Y)$=160000

由中心极限定理，$\dfrac{Y-1600}{400}$ 近似$N(0,1)$

（独立同分布下的中心极限定理）

$$P(Y>1920) = 1-P(Y\le 1920) \approx 1-\Phi\left(\frac{1920-1600}{400}\right)$$

$$= 1-\Phi(0.8) = 1-0.7881 = 0.2119$$

二、例题

【例1】一加法器同时收到20个噪声电压$V_k(k=1,2,\cdots n)$,设它们是相互独立的随机变量，且都在区间(0, 10)上服从均匀分布.记$V=\sum_{k=1}^{n}V_k$，求$P\{V>105\}$的近似值.

【解】 $V_k \sim U(a,b)$,则$E(V_k)=\frac{(a+b)}{2}$,$D(V_k)=\frac{(a+b)^2}{12}$

易知$E(V_k)=5, D(V_k)=100/12 \qquad (k=1,2,\cdots 20).$

由定理1知，$V=\sum_{k=1}^{20}V_k \overset{\text{近似地}}{\sim} N(20\times 5,\frac{100}{12}\times 20)$

（独立同分布下的中心极限定理）

于是 $P\{V>105\}=p\left\{\frac{V-20\times 5}{(\sqrt{100/12})\sqrt{20}}>\frac{105-20\times 5}{(\sqrt{100/12})\sqrt{20}}\right\}$

$=p\left\{\frac{V-20\times 5}{(\sqrt{100/12})\sqrt{20}}>0.387\right\}$

$$P\{V>105\}=p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}>\frac{105-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}\right\}$$

$$=p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}>0.387\right\}$$

$$=1-p\left\{\frac{V-20\times 5}{\left(\sqrt{100/12}\right)\sqrt{20}}\leq 0.387\right\}$$

$$\approx 1-\Phi(0.387)=0.348$$

即有 $\quad \mathrm{P}\{\mathrm{V}>105\}\approx 0.348$

【例2】 (供电问题)某车间有200台车床,在生产期间由于需要检修、调换刀具、变换位置及调换工件等常需停车. 设开工率为0.6, 并设每台车床的工作是独立的，且在开工时需电力1千瓦.

问应供应多少瓦电力就能以99.9%的概率保证该车间不会因供电不足而影响生产?

【解】 对每台车床的观察作为一次试验，每次试验是观察该台车床在某时刻是否工作, 工作的概率0.6，共进行200次独立重复试验.

用X表示在某时刻工作着的车床数,

依题意， $X \sim B(200,0.6)$,

设需N台车床工作， 现在的问题是：

求满足 $P(X\le N)\ge 0.999$ 的最小的N.

（由于每台车床在开工时需电力1千瓦，N台工作所需电力即N千瓦.）

由德莫佛-拉普拉斯极限定理（定理3）

$$\frac{X-np}{\sqrt{np(1-p)}}$$ 近似$N(0,1)$,

这里 $np=120$, $np(1-p)=48$

于是 $P(X\le N)= P(0\le X\le N)$

$$\approx \Phi\left(\frac{N-120}{\sqrt{48}}\right)-\Phi\left(\frac{-120}{\sqrt{48}}\right)$$

$$\frac{120}{\sqrt{48}}>\frac{120}{\sqrt{49}}>\frac{120}{7}>3\sigma=3$$

由3σ准则，此项为0。

$$\approx \Phi\left(\frac{N-120}{\sqrt{48}}\right)$$

由$\Phi(\frac{N-120}{\sqrt{48}}) \geq 0.999$ 查正态分布函数表得

$$\Phi(3.1) = 0.999$$

故 $\frac{N-120}{\sqrt{48}} \geq 3.1$, 从中解得$N \geq 141.5$,

即所求N=142.

也就是说, 应供应142 千瓦电力就能以99.9%的概率保证该车间不会因供电不足而影响生产.

【例3】对于一个学生而言，来参加家长会的家长人数是一个随机变量，设一个学生无家长、1名家长、2名家长来参加会议的概率分别为0.05、0.8、0.15.若学校共有400名学生，设各学生参加会议的家长数相互独立，且服从同一分布.

（1）求参加会议的家长数X超过450的概率；

（2）求有1名家长来参加会议的学生数不多于340的概率.

【解】(1)以$X_k(k=1,2\cdots,400)$记第k个学生来参加会议的家长数，则X_k的分布律为

X_k	0	1	2
p_k	0.05	0.8	0.15

易知$E(X_k)=\sum_k x_k \cdot p_k = 1.1,$

$D(X_k)=\sum_k [x_k - E(X_k)]^2 \cdot p_k = 0.19 \quad k=1,2,\cdots 400.$

而$X=\sum_{k=1}^{400} X_k$。由定理1,可知随机变量

（独立同分布下的中心极限定理）

$$X \overset{\text{近似地}}{\sim} N(400\times 1.1, 400\times 0.19)$$

即有

$$\frac{\sum_{k=1}^{400} X_k - 400\times 1.1}{\sqrt{400}\sqrt{0.19}} = \frac{X - 400\times 1.1}{\sqrt{400}\sqrt{0.19}} \overset{\text{近似地}}{\sim} N(0,1)$$

于是

$$P\{X>450\}=P\left\{\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}>\frac{450-400\times1.1}{\sqrt{400}\sqrt{0.19}}\right\}$$

$$=1-P\left\{\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}\le 1.147\right\}$$

$$\approx \mathbf{1-\Phi(1.147)=0.1257}$$

(2)以Y记有一名家长来参加会议的学生数，则

$Y\sim b(400,0.8)$，由定理3得（德莫佛-拉普拉斯极限定理）

随机变量Y $\overset{\text{近似地}}{\sim}$ **N（400×0.8,400×0.8×0.2）**

随机变量 $Y \overset{\text{近似地}}{\sim} N(400\times0.8, 400\times0.8\times0.2)$

$$P\{Y\le 340\}=P\left\{\frac{Y-400\times0.8}{\sqrt{400\times0.8\times0.2}}\le\frac{340-400\times0.8}{\sqrt{400\times0.8\times0.2}}\right\}$$

$$=P\left\{\frac{Y-400\times0.8}{\sqrt{400\times0.8\times0.2}}\le 2.5\right\}$$

$$\approx\Phi(2.5)=0.9938$$

四、小结

列维—林德伯格中心极限定理

中心极限定理

独立同分布中心极限定理
$$\begin{cases} E(X_k)=\mu,\ D(X_k)=\sigma^2 \\ \Rightarrow \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2) \end{cases}$$

棣莫弗－拉普拉斯中心极限定理
$$\begin{cases} \eta_n \sim N(n,p) \\ \Rightarrow \eta_n \overset{\text{近似地}}{\sim} N(np, np(1-p)) \end{cases}$$

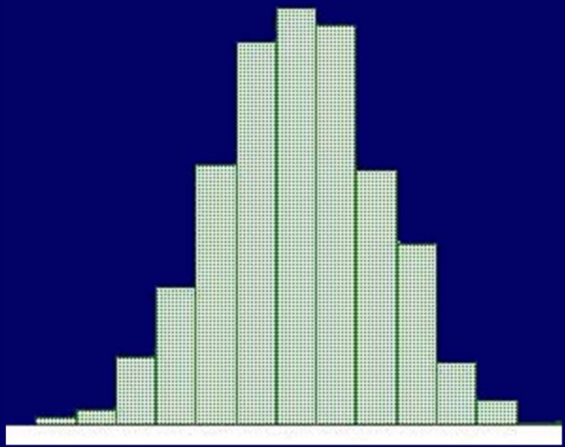
李雅普诺夫中心极限定理
$$\begin{cases} E(X_k)=\mu_k, D(x_k)=\sigma_k{}^2 \\ \Rightarrow \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(\sum_{k=1}^{n}\mu_k, B_n{}^2) \end{cases}$$

注 随机变量$X_1, X_2, \cdots$是相互独立的

这一节我们介绍了中心极限定理

中心极限定理是概率论中最著名的结果之一，它不仅提供了计算独立随机变量之和的近似概率的简单方法，而且有助于解释为什么很多自然群体的经验频率呈现出钟形曲线这一值得注意的事实.

在后面的课程中，我们还将经常用到中心极限定理.



李雅普诺夫Liapounov（1857～1918）

- 俄国著名的数学家、力学家，彼得堡数学学派的重要成员，切比雪夫的最著名最有才华的学生之一。
  - 在大学四年级时就写出了具有创见的论文

【学术成就】

- 概率论：创立了特征函数法，实现数学方法上的革命。
  - 从一个全新的角度去考察中心极限定理，比切比雪夫、马尔可夫关于中心极限定理的证明更简单而严密
  - 第一次科学地解释了为什么实际中遇到的许多随机变量近似服从正态分布
- 力学：力学中运动稳定性理论奠基人之一、常微分方程运动稳定性理论的创始人；研究旋转流体的平衡形状及其稳定性。
- 数学物理：对位势理论的研究为数学物理方法的发展开辟了新的途径，奠定了解边值问题经典方法的基础。

返回

棣莫弗De Movire（1667～1754）

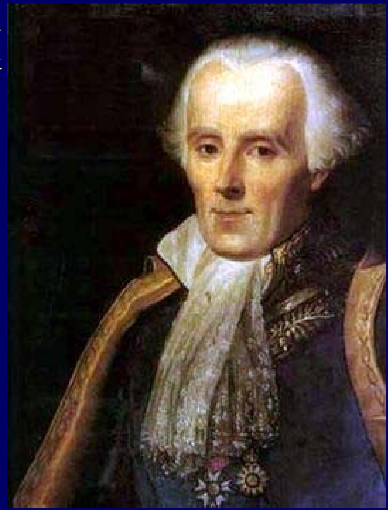
- 法国裔英国籍数学家
  - 出生于法国的一个乡村医生之家，长期生活在英国，对数学的所有贡献全是在英国做出的
  - 终生未婚，尽管在学术研究方面颇有成就，却贫困潦倒，主要以家庭教师（数学方面）为生
  - 关于棣莫弗的死有一个颇具数学色彩的神奇传说：在87岁时患上了嗜眠症，每天睡觉长达20小时。在临终前若干天，棣莫弗发现，他每天需要比前一天多睡1/4小时，那么各天睡眠时间将构成一个算术级数，当此算术级数达到24小时时，棣莫弗就长眠不醒了.

【学术成就】概率论

- 《机遇论》：较早期的概率史上有三部里程碑性质的著作，棣莫弗的《机遇论》、伯努力的《推测术》和拉普拉斯的《概率的分析理论》
- **中心极限定理**：棣莫弗第一个发现的，他在1733年发表的论文中使用正态分布去估计大量抛掷硬币出现正面次数的分布，可惜当时并未引起很大反响

返回

拉普拉斯Laplace（1749 ~ 1827）



- 法国数学家、物理学家，是天体力学的主要奠基人、天体演化学的创立者之一，他还是分析概率论的创始人，因此可以说他是应用数学的先驱。
  - 曾任拿破仑的老师，和拿破仑结下不解之缘。在数学上是大师，但在政治上却是小人物、墙头草，总是效忠于得势的一边，被人看不起，拿破仑曾讥笑他把无穷小量精神带到内阁里。.

【学术成就】

- 注意力主要集中在天体力学的研究上面，在研究过程中创造和发展了许多数学的方法，以他的名字命名的**拉普拉斯变换**、**拉普拉斯定理**和**拉普拉斯方程**，在科学技术的各个领域有着广泛的应用。
- 发表**天文学**、**数学**和**物理学**的论文270多篇，专著合计有4006多页。其中最有代表性的专著有《天体力学》（15卷16册）、《宇宙体系论》和《概率分析理论》。