

## 第五章 大数定律和中心极限定理

1. 随机变量序列的收敛性
2. 大数定律
3. 中心极限定理

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com

## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

## 引例1

设  $Y_i$  独立同分布

$$Y_i \sim B(1, p) \Rightarrow \sum_{i=1}^n Y_i \sim B(n, p)$$

$$X_n = f_n = \frac{1}{n} \sum_{i=1}^n Y_i$$

$$E(X_n) = p, D(X_n) = \frac{p(1-p)}{n}$$

$$X_n \xrightarrow{n \rightarrow \infty} p$$

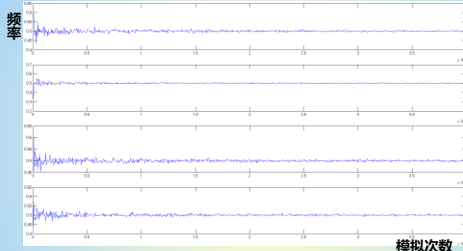
以抛硬币问题为例，  
取  $p=0.5$ ，作四次模拟  
分析：  
随着  $n$  的增大，频率  
变化趋势如何？

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

分析：随着  $n \rightarrow \infty$ ，图形趋势特点

$$\lim_{n \rightarrow \infty} X_n = p$$

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

## 定义

设  $X_1, X_2, \dots, X_n, \dots$  是随机变量序列， $X$  是一个随机变量(或为常数  $a$ )  
如果对于任意给定的正数  $\varepsilon$ ，有

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1$$

注意：  
与数  
列收  
敛的  
区别！

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

则称随机变量序列  $X_1, X_2, \dots, X_n, \dots$  依概  
率收敛于  $X$ 。

记为： $X_n \xrightarrow{P} X$  或记为  $\lim_{n \rightarrow \infty} X_n = X(P)$

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

## 引例2

设  $Y_i$  独立同分布

$$\sum_{i=1}^n Y_i \sim ?$$

$$X_n = \frac{\sum_{i=1}^n Y_i - E\left(\sum_{i=1}^n Y_i\right)}{\sqrt{D\left(\sum_{i=1}^n Y_i\right)}} \sim ?$$

$$F_n(x) = P\{X_n \leq x\} \xrightarrow{n \rightarrow \infty} ?$$

分别以各种分布为  
例，模拟其和的分布

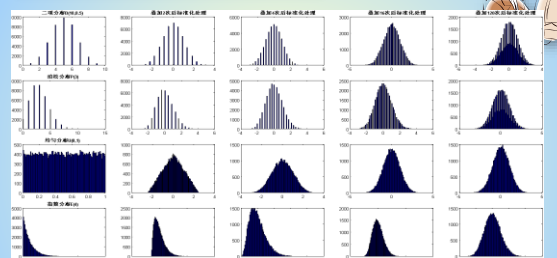
分析：  
随着  $n$  的增大，和的  
分布有何变化特点？

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

分析：随着  $n$  增大，分布函数特点？

$$\lim_{n \rightarrow \infty} F_n = \Phi(x)$$

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com

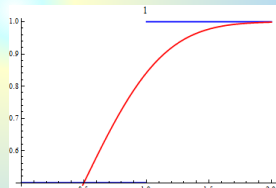


## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

设  $Y_i \sim B(1, 0.5)$ , 则  $Y_1 + Y_2 + \dots + Y_n \sim B(n, p)$ 。  
观察二项分布  $B(n, p)$  的分布函数, 红色曲线为正态分布  $N(np, np(1-p))$  的分布函数, 二者有何关系?

进一步分析:  
若  $X_n \sim B(n, p)$  进行标准化处理后, 其分布函数与  $N(0, 1)$  的分布函数有何关系?



电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

## 定义

设  $X_1, X_2, \dots, X_n, \dots$  是随机变量序列,  $X$  为随机变量,  $F_n(x)$  和  $F(x)$  分别是  $X_n$  和  $X$  的分布函数, 如果在  $F(x)$  的连续点  $x$  处均有

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

则称随机变量序列  $X_1, X_2, \dots, X_n, \dots$  依分布收敛于  $X$ 。记为:  $X_n \xrightarrow{L} X$

例如: 均匀分布之和依分布收敛于正态分布

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

$$X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{L} X \quad \text{反之不然}$$

例: (依分布收敛但不依概率收敛的反例)

设  $\Omega = \{\omega_1, \omega_2\}$ ,  $P(\omega_1) = P(\omega_2) = \frac{1}{2}$

定义随机变量:  $X(\omega_1) = -1, X(\omega_2) = 1$

则  $X$  的分布律为

$X$	-1	1
$P$	0.5	0.5

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 一. 随机变量序列的收敛性

第五章 大数定律和中心极限定理

令  $X_n(\omega) = -X(\omega)$ , 则  $X_n$  与  $X$  的分布律相同

$X$	-1	1
$P$	0.5	0.5
$X_n$	1	-1

从而分布函数相同

故  $X_n(\omega) \xrightarrow{L} X(\omega)$  即  $X_n \xrightarrow{L} X$

但对任意的  $0 < \varepsilon < 2$

$$P\{|X_n - X| > \varepsilon\} = P\{\Omega\} = 1$$

故  $X_n \not\xrightarrow{P} X$  不成立

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律

第五章 大数定律和中心极限定理

## 定理5.2.1 (切比雪夫不等式)

$$\forall \varepsilon > 0, P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2}$$

$$\text{或 } P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

[证明时先用定义展开, 再将系数变大, 最后将范围变大。]

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律

第五章 大数定律和中心极限定理

证明:

$$P(|X - \mu| \geq \varepsilon) = \int_{\{x: |x - \mu| \geq \varepsilon\}} f(x) dx \quad (\text{区域上概率, 积分})$$

$$\leq \int_{\{x: |x - \mu| \geq \varepsilon\}} \frac{|x - \mu|^2}{\varepsilon^2} f(x) dx \quad (\text{系数增大})$$

$$\leq \int_{-\infty}^{+\infty} \frac{|x - \mu|^2}{\varepsilon^2} f(x) dx \quad (\text{积分范围增大})$$

$$= \frac{1}{\varepsilon^2} \int_{-\infty}^{+\infty} (x - \mu)^2 f(x) dx$$

$$= \frac{1}{\varepsilon^2} D(X) \quad (\text{方差定义})$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律

第五章 大数定律和中心极限定理

$$\forall \varepsilon > 0, \quad P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

$$\forall \varepsilon > 0, \quad P\left\{\left|\frac{X - E(X)}{\sqrt{D(X)}}\right| < \delta\right\} \geq 1 - \frac{1}{\delta^2}$$

切比雪夫不等式只利用期望与方差就描述了随机变量的变化情况, 因此它在理论研究与实际应用中很有价值

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理

$$\forall \varepsilon > 0, \quad P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$$

例5.2.1与5.2.2:

利用切比雪夫不等式对概率进行估计。

对比例2.3.7分析:

为什么说切比雪夫不等式只能对概率作出很粗略的估计?

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理

例: 利用切比雪夫不等式粗略估计概率

设在相同条件下, 独立对某物的长度进行 $n$ 次测量, 各次测量的结果均服从正态分布  $X \sim N(a, \sigma^2)$ , 试用切比雪夫不等式估计  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  落在  $[a - 3\sigma, a + 3\sigma]$  内的概率。

解: 首先列出待求的概率表达式

$$P(a - 3\sigma < \bar{X} < a + 3\sigma) = P(-3\sigma < \bar{X} - a < 3\sigma)$$

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理

解: 首先列出待求的概率表达式

$$P(a - 3\sigma < \bar{X} < a + 3\sigma) = P(-3\sigma < \bar{X} - a < 3\sigma)$$

然后转换为切比雪夫不等式形式

$$\begin{aligned} &= P(|\bar{X} - a| < 3\sigma) = P(|\bar{X} - E(\bar{X})| < 3\sigma) \\ &\geq 1 - \frac{D(\bar{X})}{(3\sigma)^2} = 1 - \frac{\sigma^2/n}{9\sigma^2} \\ &= 1 - \frac{1}{9n} \end{aligned}$$

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理

如何根据随机变量落在区间内的概率估计  $\varepsilon$  值?

$$P(|X - E(X)| < \varepsilon) \geq a \quad (0 < a < 1)$$

方法: 这实际是上例的反问题, 主要将  $a$  与  $1 - \frac{D(X)}{\varepsilon^2}$  比较

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理

例: 设在相同条件下, 独立对某物的长度进行 $n$ 次测量, 各次测量的结果  $X_i$  均服从正态分布  $N(a, \sigma^2)$ , 问:

对该物体的长度至少要测量多少次, 才能使得用测量的平均值作为物体的长度真值的近似值, 其误差不超过 $3\sigma$ 的概率不小于0.99。

解: 由题意 $\bar{X}$ 应满足:  $P(|\bar{X} - a| \leq 3\sigma) \geq 0.99$ 

电子科技大学数学科学学院 杜冠飞 hongfdu@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



解：由题意应满足： $P\left(\left|\bar{X}-a\right|\leq 3\sigma\right)\geq 0.99$

由切比雪夫不等式  $P\left(\left|\bar{X}-a\right|\leq 3\sigma\right)\geq 1-\frac{D(\bar{X})}{\varepsilon^2}=1-\frac{1}{9n}$

从而应有  $1-\frac{1}{9n}\geq 0.99 \Rightarrow n\geq 11.11$

故至少要做2次独立重复测量，才能测量的平均值的误差不超过 $\sigma$ 的概率不小于0.99

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



## 定义

设随机变量序列  $X_1, X_2, \dots, X_n, \dots$  的每个数学期望  $E(X_i)$ ,  $i=1, 2, \dots$  均存在, 如果对于任意给定的正数  $\varepsilon$ , 有

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

则称随机变量序列  $X_1, X_2, \dots, X_n, \dots$  服从大数定律。

下面在  $X_1, \dots, X_n$  相互独立的条件下, 说明三个大数定律之间的关系:

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



• 切比雪夫不等式:  $P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2}$

• Th5.2.2(切比雪夫大数定律)

$E(X_i)$  存在,  $D(X_i) < C$ ,  $i=1, 2, 3, \dots$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right\} = 1 \quad (\varepsilon > 0)$$

$$\text{令 } X = \frac{1}{n} \sum_{i=1}^n X_i$$

$$P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right\} \geq 1 - \frac{D\left(\frac{1}{n} \sum_{i=1}^n X_i\right)}{\varepsilon^2} \geq 1 - \frac{C/n}{\varepsilon^2}$$

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



• Th5.2.2(切比雪夫大数定律)

$E(X_i)$  存在,  $D(X_i) < C$ ,  $i=1, 2, 3, \dots$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| < \varepsilon\right\} = 1 \quad (\varepsilon > 0)$$

• Th5.2.3(独立同分布大数定律)

$E(X_i) = \mu$ ,  $D(X_i) = \sigma^2$ ,  $i=1, 2, \dots$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right\} = 1 \quad (\varepsilon > 0)$$

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



• Th5.2.3(独立同分布大数定律)

$E(X_i) = \mu$ ,  $D(X_i) = \sigma^2$ ,  $i=1, 2, \dots$

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right\} = 1 \quad (\varepsilon > 0)$$

• Th5.2.5(贝努里大数定律)

$m$  为  $n$  重贝努里试验中事件  $A$  出现的次数,  $p$  为  $A$  在每次试验中发生的概率

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{m}{n} - p\right| < \varepsilon\right\} = 1 \quad (\varepsilon > 0)$$

$$X = \sum_{i=1}^n X_i = m$$

$$E(X_i) = \mu = p$$

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律

第五章 大数定律和中心极限定理



• Th5.2.4(辛钦大数定律)

设  $X_1, X_2, \dots, X_n, \dots$  是独立同分布的随机变量序列, 数学期望  $E(X_i) = \mu$ ,  $i=1, 2, \dots$  均存在, 则随机变量序列  $X_1, X_2, \dots, X_n, \dots$  服从大数定律。即对于任意的正数  $\varepsilon$ , 有

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

电子科技大学数学科学学院 杜润飞 honghefei@qq.com



## 二. 大数定律的应用

第五章 大数定律和中心极限定理

## 应用一:

1. 测量值估计: 实际工作中, 以大量测量值的平均值作为精确值的估计值→以独立同分布大数定律为理论依据

## 应用二:

2. 小概率事件原理: 概率很小的事件, 实际发生的频率也很小, 在一次试验中几乎不可能发生→实际中看作不可能事件
3. 三倍标准差原理(3σ原理): 若  $X \sim N(\mu, \sigma^2)$ , 则当X取值在  $(\mu - 3\sigma, \mu + 3\sigma)$  之外时, 看作小概率事件  $P\{|X - \mu| > 3\sigma\} = 0.0026$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律的应用

第五章 大数定律和中心极限定理

## 应用说明:

- 统计推断时, 假设检验以小概率事件原理为基础
- 很多实际问题并非理想的随机变量, 例如人的身高, 取值范围不可能为  $(-\infty, +\infty)$ , 但根据3σ原理, 考虑有限范围是有理论依据的。

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律的应用

第五章 大数定律和中心极限定理

## 应用三:

4. 矩估计的理论基础: 独立同分布大数定律可推广为辛钦大数定律 (只要求各个随机变量的期望相同即可), 可得样本矩依概率收敛于总体矩, 这在统计推断中保证了估计量的相合性

## 应用说明:

- 假设  $\{X_i\}$  是独立同分布随机变量序列, 且  $E(X_i) = E(X), i = 1, 2, \dots, n$ , 则

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E(X)$$

- 对于高阶矩有类似结论。

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律的应用

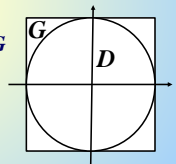
第五章 大数定律和中心极限定理

## 应用四:

5. 频率的稳定性: 试验次数充分多, 则A发生的频率  $m/n$  趋于A发生的概率  $p$ , 是计算机做随机模拟中蒙特卡罗法(Monte Carlo)的理论基础→以贝努里大数定律为理论依据

应用: 用蒙特卡罗法估计圆周率  $\pi$   
 设随机变量  $(X, Y)$  在正方形区域  $G$  内服从二维均匀分布, 内切圆  $D$  的半径为1, 根据均匀分布性质有

$$P\{(X, Y) \in D\} = \frac{S(D)}{S(G)} = \frac{\pi \times 1^2}{2 \times 2} = \frac{\pi}{4}$$



电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 二. 大数定律的应用

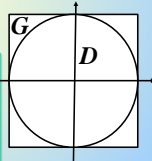
第五章 大数定律和中心极限定理

## 随机模拟方法:

$$P\{(X, Y) \in D\} = \frac{S(D)}{S(G)} = \frac{\pi \times 1^2}{2 \times 2} = \frac{\pi}{4}$$

1. 产生  $n$  组落在正方形  $G$  内的随机点  $(x, y)$ , 其中  $X \sim U(-1, 1), Y \sim U(-1, 1)$
2. 统计满足  $x^2 + y^2 \leq 1$  的随机点个数, 设为  $k$
3.  $n$  足够大时可用  $\frac{4k}{n}$  近似估算圆周率  $\pi$

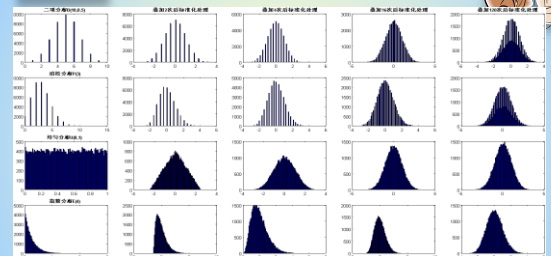
依据: 根据贝努里大数定律, 随机点落在区域  $D$  内的频率  $k/n$  依概率收敛于概率  $p = \pi/4$



电子科技大学数学科学学院 杜海飞 honghefei@qq.com

## 回顾:

第五章 大数定律和中心极限定理



分析: 随着  $n$  增大, 分布函数特点?

$$\lim_{n \rightarrow \infty} F_n = \Phi(x)$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## 正态分布与现实情形

正态分布是日常生活中最常见、常用的一种分布，如：

身高、体重、水文水位、测量数据等

实际变量往往是由诸多因素综合而成，如零件大小的测量数据 $Z$ 可能会受下列因素影响：

温度( $X_1$ )、湿度( $X_2$ )、视觉( $X_3$ )、仪器( $X_4$ )等  
故可看作  $Z=X_1+X_2+X_3+X_4+\dots$

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## 正态分布与现实情形

$$Z=X_1+X_2+X_3+X_4+\dots$$

为什么测量值 $Z$ 可看作服从正态分布的随机变量？

中心极限定理将告诉我们：(独立或弱相依) 随机变量之和的极限分布在什么条件下是正态的。

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## 定义

设 $X_1, X_2, \dots, X_n, \dots$ 是相互独立的随机变量序列，其前 $n$ 项和的标准化随机变量序列为

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\sum_{i=1}^n D(X_i)}} \quad (n=1, 2, \dots)$$

记 $Z_n$ 的分布函数为 $F_n(x)=P\{Z_n \leq x\}$ ，若

$$\lim_{n \rightarrow \infty} P\{Z_n \leq x\} = \Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt, \quad -\infty < x < +\infty$$

则称随机变量序列 $X_1, X_2, \dots, X_n, \dots$ 服从中心极限定理。

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

• 以下均设 $X_1, \dots, X_n$ 相互独立

$$\lim_{n \rightarrow \infty} P\{Z_n \leq x\} = \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n E(X_i)}{\sqrt{\sum_{i=1}^n D(X_i)}} \leq x\right\} = \Phi(x)$$

## • 独立同分布中心极限定理：

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \leq x\right\} = \Phi(x)$$

$E(X_i) = \mu, i=1, 2, \dots$   
 $D(X_i) = \sigma^2 > 0$

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## • 独立同分布中心极限定理：

$$E(X_i) = \mu$$

$$D(X_i) = \sigma^2 < \infty$$

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n\sigma^2}} \leq x\right\} = \Phi(x)$$

• 拉普拉斯中心极限定理： $Y_n \sim B(np)$ 

$$\lim_{n \rightarrow \infty} P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq x\right\} = \Phi(x)$$

$$Y_n = \sum_{i=1}^n X_i$$

$$E(X_i) = \mu = p$$

$$D(X_i) = \sigma^2 = p(1-p)$$

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

例1. 10部机器独立工作，每部停机的概率为0.2，求同时停机不多于3部的概率。

解：设10部机器中同时停机的数目为 $X$ ，则

$$X \sim B(10, 0.2), np = 2, \sqrt{np(1-p)} \approx 1.265$$

方法一：直接计算

$$P\{X \leq 3\} = \sum_{k=0}^3 C_{10}^k \cdot 0.2^k \cdot 0.8^{10-k} = 0.879$$

方法二：用中心极限定理

$$P\{X \leq 3\} \approx \Phi\left(\frac{3 - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{3-2}{1.265}\right) = 0.785$$

两种方法差异较大，原因是： $n$ 不够大！  
一般应有  $n \geq 30$

电子科技大学数学科学学院 杜海洋 hongshiduo@qq.com



## 三.中心极限定理

例2. 每颗炮弹命中飞机的概率为0.01, 求500发炮弹命中多于5发的概率。

解:  $X \sim B(500, 0.01), np = 5, \sqrt{np(1-p)} \approx 2.2$

方法一: 直接计算:

$$P\{X > 5\} = 1 - \sum_{k=0}^5 C_{500}^k \cdot 0.01^k \cdot 0.99^{500-k} \approx 0.384$$

方法二: 用中心极限定理:

$$P\{X > 5\} = P\{X \geq 6\} \approx 1 - \Phi\left(\frac{6 - np}{\sqrt{np(1-p)}}\right) = 0.327$$



## 三.中心极限定理

方法三: 用Poisson公式(查表):

$$P\{X > 5\} = P\{X \geq 6\} \approx 1 - \sum_{k=0}^5 \frac{5^k}{k!} e^{-5} \approx 0.384$$

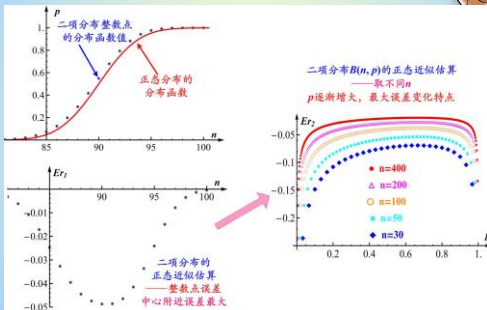
用中心极限定理差异较大, 原因是:

- 二项分布的极限分布, 当 $np$ 较小时, 用泊松分布近似计算更精确;
- 二项分布的极限分布, 当 $np$ 较大时, 可用正态分布逼近;
- 泊松分布参数 $\lambda$ 较大时可用正态分布近似计算。



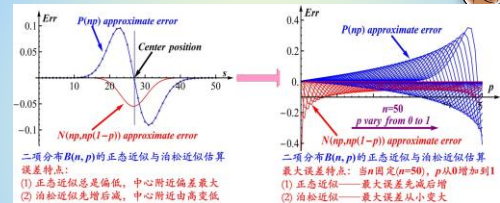
## 三.中心极限定理

扩展1: 正态分布对二项分布的近似



## 三.中心极限定理

正态分布和泊松分布对二项分布的近似计算误差



- 当 $np < 10$ , 采用泊松分布近似;
- 当 $10 < np < n - 10$ 时, 可用正态分布近似;
- 当 $n(1-p) < 10$ 时, 正态分布近似的误差也很大。



## 三.中心极限定理

扩展2: 正态分布近似估算的整数边界修正

用正态分布近似估算二项分布, 前者是连续型分布而后者是离散型分布, 当边界点为整数时容易产生分歧——如例2中 $P\{X > 5\}$ 与 $P\{X \geq 6\}$ , 同一事件的不同处理导致结果出现较大差异。

这类情形可用连续性修正方法提高估算精度并避免分歧。

为便于观察与理解修正计算方法, 以一种“极端”情形进行说明: 用 $N(1, 0.5)$ 估算 $B(2, 0.5)$ 。

## 三.中心极限定理

正态分布近似估算的整数边界修正

例: 用 $N(1, 0.5)$ 估算 $B(2, 0.5)$ 。

图中给出 $B(2, 0.5)$ 的概率直方图及相应的正态曲线, 从图可看出整数点的概率 $P\{X = k\}, k = 0, 1, 2$ 对应直方图中矩形面积, 修正计算时将其对应正态曲线在区域 $k - 0.5 \sim k + 0.5$ 之间的面积。例如,  $P\{X \leq 1\} = 0.75$ ,

- 若直接用 $\Phi(1, 0.5; 1) = 0.5$ 估算则偏差很大;
- 用 $P\{X \leq 1\} = P\{X < 1.5\} \approx \Phi(1, 0.5; 1.5) = 0.7603$ 修正后则估算偏差较小。



## 三.中心极限定理

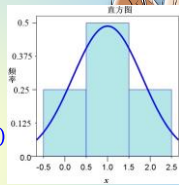
第五章 大数定律和中心极限定理

## 正态分布近似估算的整数边界修正

例：用 $N(1, 0.5)$ 估算 $B(2, 0.5)$ 。

如

$$\begin{aligned}
 0.5 &= P\{X = 1\} \\
 &\approx \Phi(1, 0.5; 1.5) - \Phi(1, 0.5; 0.5) \\
 &= 0.7603 - 0.2398 \\
 &= 0.5205
 \end{aligned}$$

表1 正态分布近似估算 $B(2, 0.5)$ 的连续性修正

$x$	-0.5	0	0.5	1	1.5	2	2.5
$P\{X=x\}$		0.25		0.5		0.25	
$P\{X \leq x\}$		0.25	0.75	1			
$\Phi(1, 0.5; x)$	0.01695	0.07805	0.2398	0.5	0.7603	0.9211	0.9831

电子科技大学数学科学学院 杜润飞 hongfdu@qq.com

## 三.中心极限定理

第五章 大数定律和中心极限定理

## 正态分布近似估算的整数边界修正

用 $N(np, np(1-p))$ 估算 $B(n, p)$ 时，边界为整数时修正方法为：根据包含的整数点向两侧扩张0.5。例如 $a, b$ 为整数时， $\{a < X\}$ 与 $\{a+1 \leq X\}$ 包含的整数边界点都是 $a+1$ ，因此修正方式为：

$$\{X = a\} \rightarrow \{a - 0.5 < X < a + 0.5\}$$

$$\{a+1 \leq X\} \cong \{a < X\} \rightarrow \{a + 0.5 < X\}$$

$$\{X \leq b-1\} \cong \{X < b\} \rightarrow \{X < b - 0.5\}$$

## 三.中心极限定理

第五章 大数定律和中心极限定理

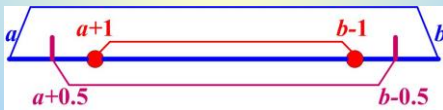
## 正态分布近似估算的整数边界修正

用 $N(np, np(1-p))$ 估算 $B(n, p)$ 时，边界为整数时修正方法为：根据包含的整数点向两侧扩张0.5。例如 $a, b$ 为整数时，修正方式为：

$$\{X = a\} \rightarrow \{a - 0.5 < X < a + 0.5\}$$

从而有

$$\begin{aligned}
 \{a+1 \leq X \leq b-1\} &\cong \{a < X < b\} \\
 &\rightarrow \{a + 0.5 < X < b - 0.5\}
 \end{aligned}$$



电子科技大学数学科学学院 杜润飞 hongfdu@qq.com

## 三.中心极限定理

第五章 大数定律和中心极限定理

## 正态分布近似估算的整数边界修正

例2. 每颗炮弹命中飞机的概率为0.01，求500发炮弹命中多于5发的概率。

解法一：二项分布直接求解 $P\{X > 5\} \approx 0.3840$ 

解法二：中心极限定理不加处理直接近似估算

$$P\{X > 5\} = 1 - P\{X \leq 5\} \approx 1 - \Phi\left(\frac{5 - np}{\sqrt{np(1-p)}}\right) = 0.5$$

解法三：中心极限定理处理后近似估算

$$P\{X > 5\} = P\{X \geq 6\} \approx 1 - \Phi\left(\frac{6 - np}{\sqrt{np(1-p)}}\right) = 0.3265$$

解法四：中心极限定理整数边界修正后近似估算

$$P\{X > 5\} = P\{X \geq 6\} = P\{X > 5.5\} \approx 1 - \Phi\left(\frac{5.5 - 5}{\sqrt{4.95}}\right) = 0.411093$$

相对误差

30.21%

相对误差

-14.97%

相对误差

7.06%

电子科技大学数学科学学院 杜润飞 hongfdu@qq.com

## 三.中心极限定理

第五章 大数定律和中心极限定理

## 如何用独立同分布中心极限定理估计概率？

方法：

首先将复杂的随机变量分解成独立同分布的随机变量之和

然后计算出随机变量之和的期望和方差

最后把变量的和看作正态分布，近似计算在某区间上的概率

电子科技大学数学科学学院 杜润飞 hongfdu@qq.com

## 三.中心极限定理

第五章 大数定律和中心极限定理

例 某工厂检验员检验一件产品每次花10秒，若有产品需重复检查则要再花10秒。假设每个产品有一半的可能性需重复检查，求在8小时内检查员检查的产品不少于1900个的概率。

思路：

首先将问题转换为求检查员检查1900个产品所花总时间 $X$ 不超过8小时的概率 $P\{X \leq 8 \cdot 3600\}$ 其次将 $X$ 转换为随机变量之和，并计算期望和方差

然后利用中心极限定理估计概率

电子科技大学数学科学学院 杜润飞 hongfdu@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

解：设检查第*i*件产品所花时间为 $X_i$  ( $i=1,2,\dots,1900$ )

则检查900件产品所花的总时间为 $K = \sum_{i=1}^n X_i$

这里， $X_i$ 是独立同分布的随机量，其中

$$X_i = \begin{cases} 10, & \text{第}i\text{件产品没有重复检查} \\ 20, & \text{第}i\text{件产品需要重复检查} \end{cases}$$

$$P\{X_i = 10\} = P\{X_i = 20\} = 0.5, \quad i = 1, 2, \dots, 1900$$

$$E(X_i) = 10 \times 0.5 + 20 \times 0.5 = 15$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

$$D(X_i) = E(X_i^2) - E(X_i)^2 = 10^2 \times 0.5 + 20^2 \times 0.5 - 15^2 = 25$$

由独立同分布中心极限定理有：

$$X = \sum_{i=1}^{1900} X_i \sim N(1900 \times 15, 1900 \times 25)$$

从而所求概率为：

$$P\{X \leq 8 \times 3600\}$$

$$\approx \Phi\left(\frac{8 \times 3600 - 1900 \times 15}{\sqrt{1900 \times 25}}\right) = \Phi\left(\frac{6}{\sqrt{19}}\right)$$

$$= \Phi(1.37649) = 0.9162$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## 棣莫弗——拉普拉斯中心极限定理的几种用法

例、设某种工艺需要合格的产品100个，该产品的合格率为96%，问需要采购多少个这种产品，才能有95%以上的把握保证合格数够用？

解：设采购*n*件产品，其中合格品数为*X*，要保证合格品数够用*X*必须满足  $X \geq 100$

因此问题归结为求最小的正整数*n*，使得

$$P(X \geq 100) \geq 0.95$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

因此问题归结为求最小的正整数*n*，使得

$$P(X \geq 100) \geq 0.95$$

由于 $X \sim B(n, 0.96)$ ，由棣莫弗——拉普拉斯中心极限定理得

$$X \sim N(np, np(1-p)) \text{ 即 } N(0.96n, 0.96 \times 0.04n)$$

从而有  $P(X \geq 100) = 1 - P(X < 100)$

$$\approx 1 - \Phi\left(\frac{100 - 0.96n}{\sqrt{0.96 \times 0.04n}}\right)$$

$$= \Phi\left(\frac{0.96n - 100}{\sqrt{0.96 \times 0.04n}}\right) \geq 0.95$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

$$\Phi\left(\frac{0.96n - 100}{\sqrt{0.96 \times 0.04n}}\right) \geq 0.95$$

$$\text{查表得: } \frac{0.96n - 100}{\sqrt{0.96 \times 0.04n}} \geq 1.64$$

$$\text{化简得: } n - 0.3348\sqrt{n} - 104.1667 \geq 0$$

$$\text{求解得: } \sqrt{n} \geq 10.3750 \text{ 或 } \sqrt{n} \leq -10.0402 (\text{舍去})$$

$$\text{也即 } n \geq 107.6406$$

故 至少需要采购108个产品才能满足要求

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



## 三.中心极限定理

第五章 大数定律和中心极限定理

## 扩展思考：是否需要考虑上限？

例、设某种工艺需要某合格的产品100个，该产品的合格率为96%，问需要采购多少个这种产品，才能有95%以上的把握保证合格品数够用？

解：设采购*n*件产品，其中合格品数为*X*，要保证合格品数够用，则要求最小的正整数*n*，使得： $P\{X \geq 100\} \geq 0.95$

思考：为保证准确度，是否应该采用  $P\{100 \leq X \leq n\} \geq 0.95$

$$P\{100 \leq X \leq n\} \approx \Phi\left(\frac{n - 0.96n}{\sqrt{0.96 \times 0.04n}}\right) - \Phi\left(\frac{100 - 0.96n}{\sqrt{0.96 \times 0.04n}}\right)$$

由于上下限并非关于中心对称，求解困难！——注意到

$$\frac{n - 0.96n}{\sqrt{0.96 \times 0.04n}} = \frac{0.04}{\sqrt{0.96 \times 0.04}} \sqrt{n} \xrightarrow{n=100} \sqrt{n} \approx 0.204 \sqrt{n} \xrightarrow{n=100} 2.04$$

$$\Phi(2.04) = 0.9793 \approx 1$$

电子科技大学数学科学学院 杜海飞 honghefei@qq.com



### 三.中心极限定理

第五章 大数定律和中心极限定理



#### 定理用法:

在近似计算公式中

$$\begin{aligned}\beta &= P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = P\left(p - \varepsilon < \frac{n_A}{n} < p + \varepsilon\right) \\ &= P(np - n\varepsilon < n_A < np + n\varepsilon) = \Phi\left(\frac{n\varepsilon}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{-n\varepsilon}{\sqrt{np(1-p)}}\right) \\ &= 2\Phi\left(\varepsilon \sqrt{\frac{n}{p(1-p)}}\right) - 1 = \beta\end{aligned}$$

有四个相互联系的指标

试验总次数, 事件A出现的概率, 误差界 $\varepsilon$

用频率估计概率所产生的误差的可靠性

只要知道其中任意三量, 就可以求出另外一个

电子科技大学数学科学学院 杜海飞 honghefei@qq.com

