

## 第五章

# 大数定律与中心极限定理

## § 2、中心极限定理

教学时数：70分钟

教学目的：

1. 了解中心极限定理的实际背景；
2. 理解两个基本的中心极限定理,并能正确解决相关问题.

# 一、实际背景：正态分布的广泛存在性。

对大量随机现象观察得到的另一个重要结果是：正态分布极为常见。

观察表明：如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大，则这种量一般都服从或近似服从正态分布。

## 二、数学描述

$r. v. X$

独立  $r.v.s. X_1, X_2, \dots$

观察表明：如果一个量  
是由大量相互独立的随机因素  
的影响所造成，  
而每一个别因素在总影响中所  
起的作用不大。

$$X = X_1 + X_2 + \dots$$

$X_1, X_2, \dots$  同分布，不妨  
设  $EX_i = \mu, DX_i = \sigma^2$

则这种量一般都服从或  
近似服从正态分布。

当  $n \rightarrow \infty$  时， $X = X_1 + X_2 + \dots + X_n$   
的分布函数的极限是正态分  
布分布函数。

“当 $n \rightarrow \infty$ 时,  $X = X_1 + X_2 + \dots + X_n$ 的分布函数的极限是正态分布的分布函数” 的数学表示?

$$n \rightarrow \infty \text{ 时, } X = X_1 + X_2 + \dots + X_n \xrightarrow{\sim} N(n\mu, n\sigma^2)$$

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{\sim} N(0, 1)$$

$\downarrow$   
 $\infty$        $\downarrow$   
 $\infty$

即

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

### 三、几个基本的中心极限定理

**定理1**（Lévy中心极限定理（独立同分布下的Lindberg中心极限定理））

设 $X_1, X_2, \dots$ 是独立同分布的随机变量序列，且 $E(X_i) = \mu$ ， $D(X_i) = \sigma^2$ ， $i=1, 2, \dots$ ，则

$$\lim_{n \rightarrow \infty} P\left\{\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

## 定义:

设随机变量序列 $\{X_n\}$ 和随机变量 $X$ 的分布函数分别为 $\{F_n(x)\}$ 和 $F(x)$ ,  $n=1,2,\dots$ , 若对 $F(x)$ 的一切连续点 $x$ , 有:  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ , 则称 $X_n$ 依分布收敛于 $X$ 。

## 定理1的重述:

设 $X_1, X_2, \dots$ 是独立同分布的随机变量序列, 且 $E(X_i) = \mu$ ,  $D(X_i) = \sigma^2$ ,  $i=1,2,\dots$ ,  $X$ 服从标准正态分布, 则  $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$  依分布收敛于 $X$ 。

$$\lim_{n \rightarrow \infty} P\left\{ \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

下面考虑贝努里试验序列.

设  $Y_n$  是  $n$  重贝努里试验中事件  $A$  发生的次数,  $p$  是事件  $A$  发生的概率,

引入 
$$X_i = \begin{cases} 1, & \text{如第 } i \text{ 次试验 } A \text{ 发生} \\ 0, & \text{如第 } i \text{ 次试验 } A \text{ 不发生} \end{cases} \quad i=1,2,\dots,n$$

则  $X_1, X_2, \dots, X_n, \dots$  i.i.d. 且  $EX_i = p$ ,  $DX_i = p(1-p)$ , 而

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

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

即有



## 定理2 (棣莫佛—拉普拉斯定理)

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

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

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  $Y_n$  的分布近似正态分布  $N(np, np(1-p))$ .

棣莫佛(Abraham de Moivre, 1667-1754)

1733年完成对二项式 $(a+b)^n$

( $a=b$ 情形) 展开式中的项求和的逼近问题。有关研究结果作为《机会学说》二、三版附录出版, 首次提出二项分布的正态逼近的概念, 发现正态曲线。



Abraham de Moivre (1667-1754)

拉普拉斯1810年在一篇论文中给出正态规律的一个新的理论推导——实际上是对De Moivre结果的推广。有关结果及证明后来收入拉普拉斯1812年的著作《分析概率论》。



Pierre-Simon Laplace (1749 -1827 )

高斯（1809）：测量误差服从正态分布。



之后，人们发现，正态分布极为常见。

例如：弗里德里希.贝塞尔对几百颗星球作了三组观测，并比较了按照正态规律在给定范围内的理论误差值和实际值。这个对比表明它们非常接近一致。

**观察表明：**如果一个量是由大量相互独立的随机因素的影响所造成，而每一个别因素在总影响中所起的作用不大。则这种量一般都服从或近似服从正态分布。

**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.

分析：设第 $i$ 只元件的寿命为 $X_i$ ,  $i=1,2,\dots,16$

由题给条件知，诸 $X_i$ 独立同分布，

$$E(X_i)=100, D(X_i)=10000$$

16只元件的寿命的总和为

依题意，所求为 $P(Y>1920)$

$$Y = \sum_{k=1}^{16} X_k$$

解：设第 $i$ 只元件的寿命为 $X_i, i=1,2, \dots, 16$

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

16只元件的寿命的总和为  $Y = \sum_{k=1}^{16} X_k$   
依题意, 所求为 $P(Y>1920)$ 。

由于 $E(Y)=1600, D(Y)=160000$

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

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

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

**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?

解: 对每台车床的观察作为一次试验, 每次试验观察该台车床在某时刻是否工作, 工作的概率为0.6, 共进行200次试验. 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 $N$ 千瓦电力, 现在的问题是:

求满足  $P(X \leq N) \geq 0.999$  的最小的 $N$ .

## 由德莫佛-拉普拉斯极限定理

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

这里  $np=120$ ,  
 $np(1-p)=48$

于是  $P(X \leq N) = P(0 \leq X \leq N)$

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

由  $\Phi\left(\frac{N-120}{\sqrt{48}}\right) \geq 0.999$ ,

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  $N=142$ .

由  $3\sigma$  准则,  
此项为0。

故供应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。



**\*练1:** 今要设置一项保险: 一辆自行车年交保费2元, 若自行车丢失, 保险公司赔偿200元. 若在一年内自行车丢失的概率为0.001, 问至少要有多少辆自行车投保才能以不小于0.9的概率保证这一保险不亏本?

解: 设有 $n$ 辆自行车投保,  $Y$ 表示一年内 $n$ 辆自行车中丢失的数量。则 $Y \sim b(n, 0.001)$ , 问题归结为 $n$ 至少为多少时,  $P\{2n - 200Y \geq 0\} \geq 0.9$ , 即  $P\{Y \leq 0.01n\} \geq 0.9$ . 由

$$\begin{aligned} P\{Y \leq 0.01n\} &= P\{0 \leq Y \leq 0.01n\} \\ &= P\left\{ \frac{-0.001n}{\sqrt{0.000999n}} \leq \frac{Y - 0.001n}{\sqrt{0.000999n}} \leq \frac{0.01n - 0.001n}{\sqrt{0.000999n}} \right\} \\ &= P\left\{ \frac{-0.001n}{\sqrt{0.000999n}} \leq \frac{Y - 0.001n}{\sqrt{0.000999n}} \leq \frac{0.009n}{\sqrt{0.000999n}} \right\} \end{aligned}$$

$$\begin{aligned}
& P\{Y \leq 0.01n\} = P\{0 \leq Y \leq 0.01n\} \\
&= P\left\{ \frac{-0.001n}{\sqrt{0.000999n}} \leq \frac{Y - 0.001n}{\sqrt{0.000999n}} \leq \frac{0.01n - 0.001n}{\sqrt{0.000999n}} \right\} \\
&= P\left\{ \frac{-0.001n}{\sqrt{0.000999n}} \leq \frac{Y - 0.001n}{\sqrt{0.000999n}} \leq \frac{0.009n}{\sqrt{0.000999n}} \right\} \\
&\approx \Phi\left(\frac{0.009n}{\sqrt{0.000999n}}\right) - \Phi\left(\frac{-0.001n}{\sqrt{0.000999n}}\right) \\
&= \Phi\left(\frac{0.009n}{\sqrt{0.000999n}}\right) - 1 + \Phi\left(\frac{0.001n}{\sqrt{0.000999n}}\right) \geq 0.9
\end{aligned}$$

所以要使上式成立

必须  $\Phi\left(\frac{0.001n}{\sqrt{0.000999n}}\right) > 0.9$

查表得  $\frac{0.001n}{\sqrt{0.000999n}} > 1.29$  解此不等式得  $n \geq 1663$ .

由3σ准则，  
当n>13时  
此项为1。

**练2:** 一加法器同时收到20个噪声电压 $V_k(k=1,2,\dots,20)$ , 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  $V = \sum_{k=1}^{20} V_k$ , 求 $P\{V > 105\}$  的近似值. (教材p. 197:5.3)

**解:** 易知 $E(V_k)=5$ ,  $D(V_k)=100/12(k=1, 2, \dots, 20)$ 。

由定理1, 随机变量

$$\frac{\sum_{k=1}^n V_k - 20 \times 5}{\sqrt{100/12} \cdot \sqrt{20}} = \frac{V - 20 \times 5}{\sqrt{100/12} \cdot \sqrt{20}} \quad \text{近似服从正态分布 } N(0, 1)$$

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

$$\begin{aligned}\text{于是 } P\{V > 105\} &= P\left\{\frac{V - 20 \times 5}{\sqrt{100/12} \cdot \sqrt{20}} > \frac{105 - 20 \times 5}{\sqrt{100/12} \cdot \sqrt{20}}\right\} \\&= P\left\{\frac{V - 20 \times 5}{\sqrt{100/20} \cdot \sqrt{20}} > 0.387\right\} \\&= 1 - P\left\{\frac{V - 20 \times 5}{\sqrt{100/20} \cdot \sqrt{20}} \leq 0.387\right\} \\&\approx 1 - \Phi(0.387) \approx 0.348\end{aligned}$$

即有  $P\{V > 105\} \approx 0.348$ 。

定理1要求随机变量序列 $\{X_n\}$ 独立同分布，若去掉同分布的条件，需要附加什么条件才能使 $\{X_n\}$ 服从中心极限定理（即使其部分和序列依分布收敛于正态变量）呢？换句话说此时怎样刻画“每一项的作用都不太大”呢？对此，Lindberg1922年证明了如下的

### 定理3(林德贝格Lindberg中心极限定理)

设随机变量 $X_1, X_2, \dots, X_n, \dots$ ，相互独立， $E(X_k)=\mu_k$ ， $D(X_k)=\sigma_k^2 \neq 0$  ( $k=1, 2, \dots$ )， $B_n^2 = \sum_{k=1}^n \sigma_k^2$

若 $\{X_n\}$ 满足林德贝格条件：对于任意的  $\tau > 0$ , 有

$$\lim_{n \rightarrow \infty} \frac{1}{B_n^2} \sum_{k=1}^n \int_{|x - \mu_k| > \tau B_n} (x - \mu_k)^2 dF_k(x) = 0$$

则对于任意实数 $x$ , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

### \*定理4 (李雅普诺夫Liapunov定理)

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 它们具有数学期望和方差,  $E(X_k) = \mu_k, D(X_k) = \sigma_k^2 \neq 0$  ( $k=1, 2, \dots$ ), 记

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

若存在正数 $\sigma$ , 使得当 $n \rightarrow \infty$ 时,

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

则对任意实数 $x$ , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt$$

注释：(1) 定理4表明，在定理的条件下，随机变量，

$$Z_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n}$$

当n很大时，近似地服从正态分布 $N(0, 1)$ 。由此，  
当n很大时，

$$\sum_{k=1}^n X_k = B_n Z_n + \sum_{k=1}^n \mu_k$$

近似地服从正态分布  $N\left(\sum_{k=1}^n \mu_k, B_n^2\right)$

(2) 同时定理也提供了大量独立随机变量之和有关的事件概率的近似计算方法.