第二节　　中心极限定理

一、问题的引入

二、基本定理

三、典型例题

四、小结

一、问题的引入

实例: **考察射击命中点与靶心距离的偏差.**

这种偏差是大量微小的偶然因素造成的微小误差的总和, 这些因素包括: 瞄准误差、测量误差、子弹制造过程方面 (如外形、重量等) 的误差以及射击时武器的振动、气象因素(如风速、风向、能见度、温度等) 的作用, 所有这些不同因素所引起的微小误差是相互独立的, 并且它们中每一个对总和产生的影响不大.

问题: 某个随机变量是由大量相互独立且均匀小的随机变量相加而成的, 研究其概率分布情况.

二、基本定理

定理四（独立同分布的中心极限定理）

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：$E(X_k) = \mu$, $D(X_k) = \sigma^2 > 0 \ (k = 1, 2, \cdots)$, 则随机变量之和的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 x 满足

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

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

定理四表明:

当 $n \to \infty$, 随机变量序列 Y_n 的分布函数收敛于标准正态分布的分布函数.

定理五(李雅普诺夫定理)

李雅普诺夫

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:

$$E(X_k) = \mu_k, \quad D(X_k) = \sigma_k^2 \neq 0 \ (k = 1,2,\cdots),$$

记

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

若存在正数 δ, 使得当 $n \to \infty$ 时,

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

则随机变量之和的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 x 满足

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

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

定理五表明:

无论各个随机变量 $X_1, X_2, \cdots, X_n, \cdots$服从什么分布, 只要满足定理的条件, 那么它们的和 $\sum_{k=1}^{n} X_k$ 当 n 很大时,近似地服从正态分布.

(如实例中射击偏差服从正态分布)

下面介绍的定理六是定理四的特殊情况.

定理六(德莫佛—拉普拉斯定理)

德莫佛 拉普拉斯

设随机变量 $\eta_n\ (n=1,2,\cdots)$ 服从参数为 n,p $(0<p<1)$的二项分布, 则对于任意 x, 恒有

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

证明 根据第四章第二节例题可知 $\eta_n = \sum_{k=1}^{n} X_k$,

其中 $X_1, X_2, \cdots, X_n$ 是相互独立的、服从同一 (0—1) 分布的随机变量, 分布律为

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

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

根据定理四得

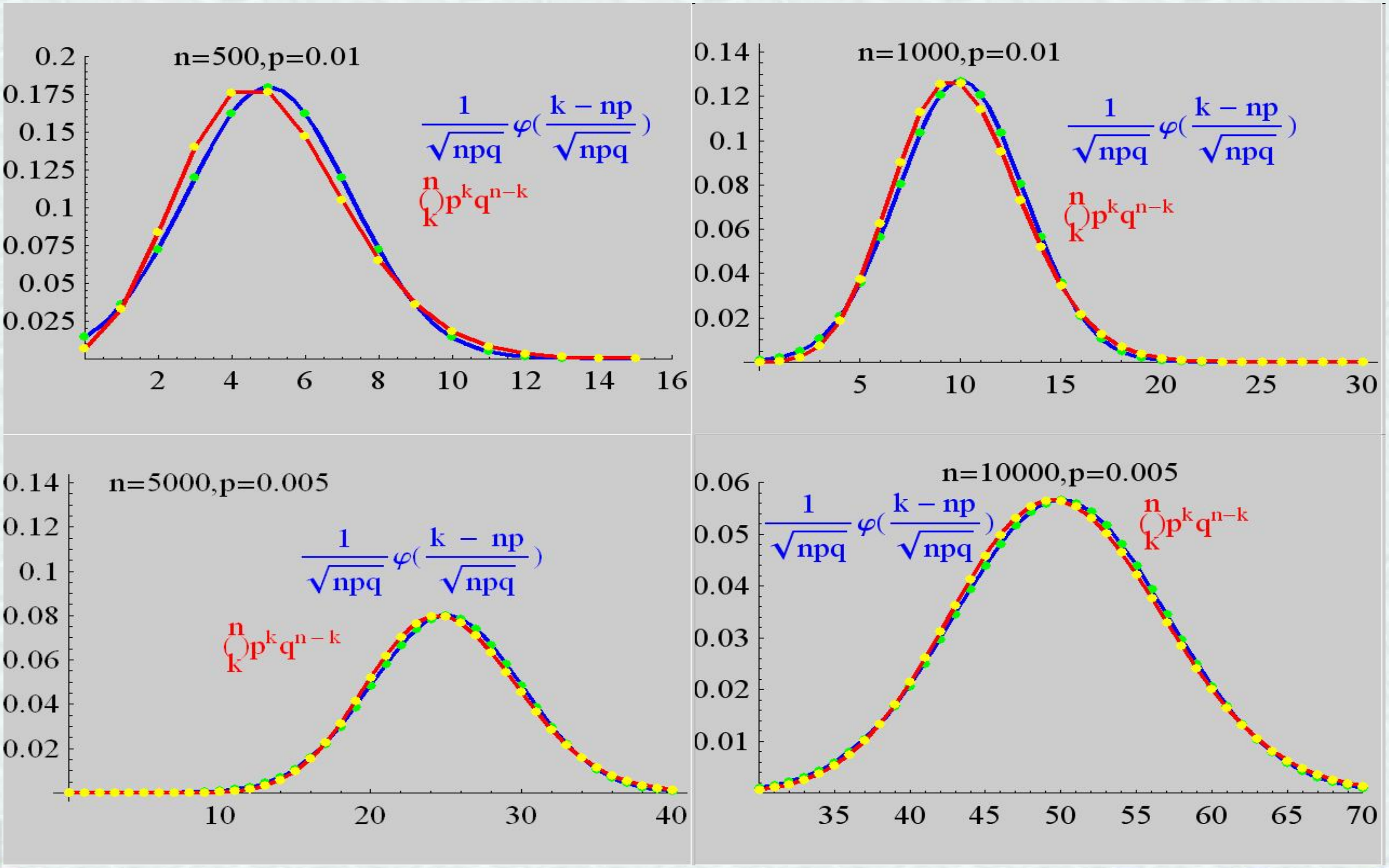
$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \lim_{n\to\infty} P\left\{\frac{\sum_{k=1}^{n} X_k - np}{\sqrt{np(1-p)}} \le x\right\}$$

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

定理六表明:

正态分布是二项分布的极限分布, 当n充分大时, 可以利用该定理来计算二项分布的概率.

下面的图形表明:正态分布是二项分布的逼近.



三、典型例题

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

解 $\because E(V_k)=5,\quad D(V_k)=\dfrac{100}{12}\ (k=1,2,\cdots,20).$

由定理四, 随机变量 Z 近似服从正态分布 $N(0,1)$,

其中 $Z=\dfrac{\sum_{k=1}^{20}V_k-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}=\dfrac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}$

$$\therefore P\{V>105\}=P\left\{\frac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}>\frac{105-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}\right\}$$

$$=P\left\{\frac{V-20\times 5}{\sqrt{\frac{100}{12}}\sqrt{20}}>0.387\right\}=1-P\left\{\frac{V-100}{\sqrt{\frac{100}{12}}\sqrt{20}}\le 0.387\right\}$$

$$\approx 1-\int_{-\infty}^{0.387}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=1-\Phi(0.387)=0.348.$$

例2　一船舶在某海区航行, 已知每遭受一次海浪的冲击, 纵摇角大于 3° 的概率为1/3, 若船舶遭受了90 000次波浪冲击, 问其中有29 500～30 500次纵摇角大于 3° 的概率是多少?

解　将船舶每遭受一次海浪的冲击看作一次试验,

并假设各次试验是独立的,

在90 000次波浪冲击中纵摇角大于 3° 的次数为 X,

则 X 是一个随机变量, 且 $X \sim b(90000, \frac{1}{3})$.

分布律为 $P\{X=k\}=\binom{90000}{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90000-k}$,

$k=1,\cdots,90000.$

所求概率为

$$P\{29500<X\le 30500\}=\sum_{k=29501}^{30500}\binom{90000}{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90000-k}.$$

直接计算很麻烦，利用**德莫佛－拉普拉斯定理**

$$P\{29500<X\le 30500\}$$

$$=P\left\{\frac{29500-np}{\sqrt{np(1-p)}}<\frac{X-np}{\sqrt{np(1-p)}}\le\frac{30500-np}{\sqrt{np(1-p)}}\right\}$$

$$\approx \int_{\frac{29500-np}{\sqrt{np(1-p)}}}^{\frac{30500-np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

$$= \Phi\left(\frac{30500-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29500-np}{\sqrt{np(1-p)}}\right)$$

$$\because n = 90000, \quad p = \frac{1}{3},$$

$$\therefore P\{29500 < X \le 30500\} \approx \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right)$$

$$= 0.9995.$$

例3 某保险公司的老年人寿保险有1万人参加,每人每年交200元. 若老人在该年内死亡,公司付给家属1万元. 设老年人死亡率为0.017,试求保险公司在一年内的这项保险中亏本的概率.

解 设 X 为一年中投保老人的死亡数,

则 $X \sim B(n, p)$,

其中 $n = 10000$, $p = 0.017$,

由德莫佛—拉普拉斯定理知,

保险公司亏本的概率

$$P\{10000X > 10000\times 200\} = P\{X > 200\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > \frac{200-np}{\sqrt{np(1-p)}}\right\}$$

$$= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > 2.321\right\}$$

$$\approx 1-\Phi(2.321) \approx 0.01 .$$

例4 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 $E(X_k)=1.1,\ D(X_k)=0.19,\quad (k=1,2,\cdots,400)$

而 $X=\sum_{k=1}^{400}X_k$, 根据**独立同分布的中心极限定理,**

随机变量 $$\frac{\sum_{k=1}^{400}X_k-400\times1.1}{\sqrt{400}\sqrt{0.19}}=\frac{X-400\times1.1}{\sqrt{400}\sqrt{0.19}}$$

近似服从正态分布 $N(0,1)$,

于是 $P\{X > 450\}$

$$= P\left\{\frac{X - 400\times 1.1}{\sqrt{400}\sqrt{0.19}} > \frac{450 - 400\times 1.1}{\sqrt{400}\sqrt{0.19}}\right\}$$

$$= 1 - P\left\{\frac{X - 400\times 1.1}{\sqrt{400}\sqrt{0.19}} \le 1.147\right\}$$

$$\approx 1 - \Phi(1.147) = 0.1357;$$

(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**德莫佛—拉普拉斯定理**知,

$$P\{X \le 350\}$$

$$= P\left\{\frac{Y-400\times 0.8}{\sqrt{400\times 0.8\times 0.2}} \le \frac{340-400\times 0.8}{\sqrt{400\times 0.8\times 0.2}}\right\}$$

$$= P\left\{\frac{Y-400\times 0.8}{\sqrt{400\times 0.8\times 0.2}} \le 2.5\right\} \approx \Phi(2.5) = 0.9938\,.$$

例5 设随机变量 $X_1, X_2, \cdots, X_n$ 相互独立, 且 X_i 在区间 $(-1, 1)$ 上服从均匀分布 $(i = 1, 2, \cdots, n)$, 试证当 n 充分大时, 随机变量 $Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 近似服从正态分布, 并指出其分布参数.

证 记 $Y_i = X_i^2$, $(i = 1,2,\cdots,n)$

$$E(Y_i) = E(X_i^2) = D(X_i) = \frac{1}{3},$$

$$D(Y_i) = E(Y_i^2) - [E(Y_i)]^2 = E(X_i^4) - [E(Y_i)]^2.$$

因为 $E(X_i^4)=\int_{-1}^{1}x_i^4\cdot\frac{1}{2}\mathrm{d}x_i=\frac{1}{5},$

所以 $D(Y_i)=\frac{1}{5}-\left(\frac{1}{3}\right)^2=\frac{4}{45},$

因为 $X_1,X_2,\cdots,X_n$ 相互独立,

所以 $Y_1,Y_2,\cdots,Y_n$ 相互独立.

根据独立同分布的中心极限定理,

$$n \cdot Z_n = \sum_{i=1}^{n} X_i^2 = \sum_{i=1}^{n} Y_i$$

近似服从正态分布 $N\left(\frac{n}{3}, \frac{4n}{45}\right)$,

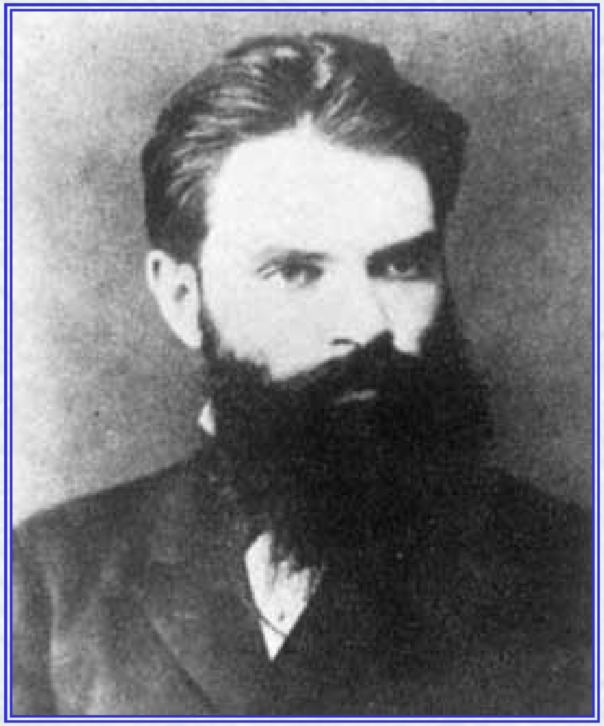
故 Z 近似地服从正态分布 $N\left(\frac{1}{3}, \frac{4}{45n}\right)$.

四、小结

三个中心极限定理
- 独立同分布的中心极限定理
- 李雅普诺夫定理
- 德莫佛—拉普拉斯定理

中心极限定理表明, 在相当一般的条件下, 当独立随机变量的个数增加时, 其和的分布趋于正态分布.

李雅普诺夫资料

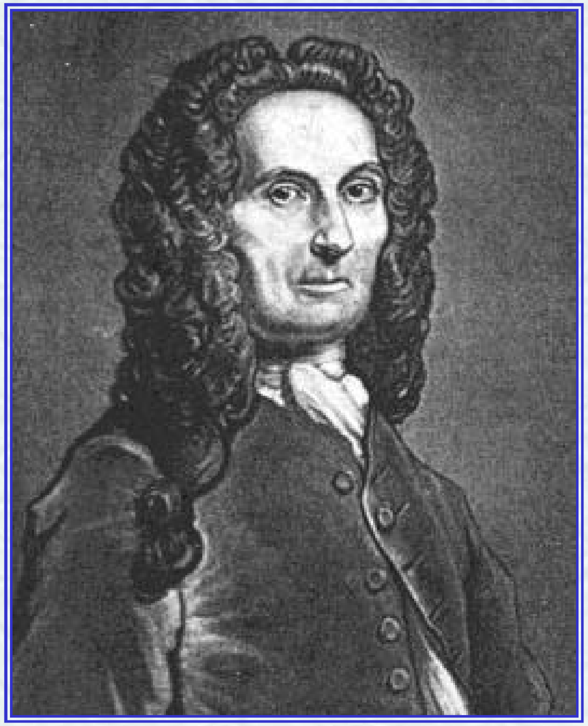


Aleksandr Mikhailovich Lyapunov

Born: **6 Jun. 1857 in Yaroslavl, Russia**

Died: **3 Nov. 1918 in Odessa, Russia**

德莫佛资料



Abraham de Moivre

Born: 26 May. 1667 in Vitry (near Paris), France

Died: 27 Nov. 1754 in London, England

拉普拉斯资料

Pierre-Simon Laplace

Born: 23 Mar. 1749 in Beaumont-en-Auge, Normandy, France

Died: 5 Mar. 1827 in Paris, France