西北工業大學
NORTHWESTERN POLYTECHNICAL UNIVERSITY

概率论与数理统计

第二节 中心极限定理

- 一、问题的提出
- 二、中心极限定理

一、问题的提出

由上一节大数定律,我们得知满足一定条件的随机变量序列的算术平均值依概率收敛, 但我们无法得知其收敛的速度, 本节的中心极限定理可以解决这个问题.

在实际中, 人们发现 n 个相互独立同分布的随机变量之和的分布近似于正态分布, 并且 n 越大,近似程度越好.

二、中心极限定理

定理4.7　林德贝格-列维中心极限定理

设随机变量$X_1, X_2, \ldots, X_n$相互独立, 服从同一分布, 且具有数学期望与方差

$$E(X_i) = \mu,\quad D(X_i) = \sigma^2 \neq 0 \quad (i=1, 2, \ldots, n)$$

则随机变量

$$Y_n^* = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$$

的分布函数$F_n(x)$ 对于任意 x 满足

目　上　下　回　停　续

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\{Y_n^* \le x\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

注 1° 当充分大时,随机变量Y_n^*近似服从标准正态分布$N(0,1)$,记为$Y_n^* \sim AN(0,1)$. n越大,近似程度越好.

2° $$Y_n = \sum_{i=1}^{n} X_i \sim AN(n\mu, n\sigma^2)$$

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \sim AN\left(\mu, \frac{\sigma^2}{n}\right)$$

3° 定理4.7表明n个相互独立同分布的随机变量的和近似服从正态分布.

定理4.8　棣莫佛-拉普拉斯定理

设随机变量Y_n服从二项分布$B(n, p)$, 则其标准化随机变量

$$Y_n^* = \frac{Y_n - np}{\sqrt{np(1-p)}}$$

的分布函数的极限为

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

证　令

$$X_i=\begin{cases}1, & \text{第}i\text{次试验}A\text{发生}\\ 0, & \text{第}i\text{次试验}A\text{不发生}\end{cases}\qquad (i=1,2,\cdots,n)$$

$X_1, X_2,\ldots, X_n$独立, 同时服从$B(1, p)$分布, 且

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

由于 $E(X_i)=p$, $D(X_i)=p(1-p)$ $(i=1, 2,\ldots, n)$,

目 上 下 回 停 续

由定理4.7得

$$\lim_{n\to\infty} P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \le x\right\}$$

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

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t$$

证毕.

注 1° **定理4.8**表明正态分布是二项分布的极限分布也称为“二项分布的正态近似”.

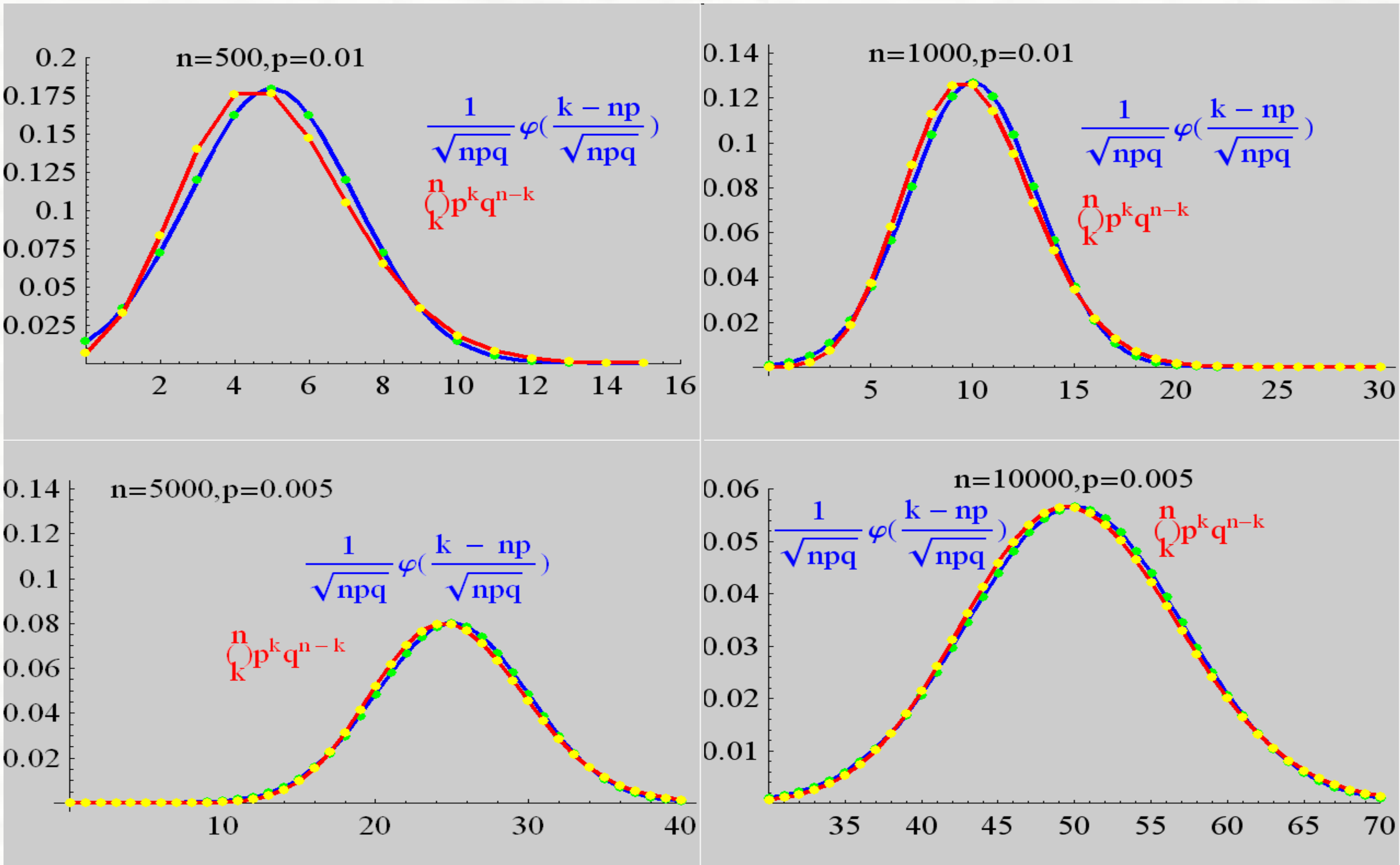
2° 与“二项分布的泊松近似”相比较, 两种近似都要求n很大.

3° 实际应用中当n很大时,

(1) 如果p很小而np不太大时，采用泊松近似;

(2) 如果 $np \geq 5$ 和 $n(1-p) \geq 5$ 同时成立时, 采用正态近似.

下面的图形表明:正态分布是二项分布的逼近.



内容小结

独立同分布情形

中心极限定理
- 林德贝格–列维中心极限定理
- 棣莫佛–拉普拉斯定理

二项分布的正态近似

目 上 下 回 停 续


再见

备用题

例1-1 设随机变量$X_1, X_2, \ldots, X_n$相互独立，且 X_i 在区间$(-1, 1)$ 上服从均匀分布$(i=1, 2, \ldots, n)$，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

$$Y_i = X_i^2, \quad (i = 1, 2, \cdots, n) \quad E(Y_i) = E(X_i^2) = D(X_i)$$

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

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$X_1, X_2,\ldots, X_n$相互独立, 所以$Y_1, Y_2,\ldots,Y_n$相互独立, 根据**定理4.8**

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

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

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

例1-2 某汽车销售点每天出售汽车数服从参数为2的泊松分布. 若一年365天都经营汽车销售, 且每天出售的汽车是相互独立的, 求一年中售出700辆以上汽车的概率.

解 记X_i为第i天出售的汽车数量,

$Y = X_1 + X_2 + \cdots + X_{365}$为一年的总销量.

由$E(X_i) = D(X_i) = 2$, 知$E(Y) = D(Y) = 730$.

利用**林德贝格-列维中心极限定理**, 可得

$$P(Y > 700) = 1 - P(Y \le 700) \approx 1 - \Phi\left(\frac{700 - 730}{\sqrt{730}}\right)$$

$$= 1 - \Phi(-1.11) = 0.8665.$$

则一年售出700辆以上汽车的概率近似为0.8665.

例1-3 某餐厅每天接待400名顾客, 设每位顾的消费额(元)服从(20, 100)上的均匀分布, 且顾客的消费额是相互独立的. 试求:

(1)该餐厅每天的平均营业额;

(2)该餐8厅每天的营业额在平均营业额±760元的概率.

解　设X_i为第i位顾客的消费额,

$X_i \sim U(20, 100)$. 所以 $E(X_i) = 60$,

$D(X_i) = 1600/3$.

而该餐厅每天的营业额为$Y = \sum_{i=1}^{400} X_i$.

(1)该餐厅每天的营业额为

$$E(Y)=\sum_{i=1}^{400}E(X_i)=400\times 60=24000$$

(2)利用**林德贝格-列维中心极限定理**, 可得

$$P(-760<Y-24000<760)\approx 2\Phi\left(\frac{760}{\sqrt{400\times 1600/3}}\right)-1$$

$$=2\Phi(1.645)-1=0.90$$

这表明:该餐厅每天的营业额在23240到24760之间的概率近似为0.90.

目 上 下 回 停 续

例1-4 某人钓鱼平均每次钓到2kg, 方差2.25kg^2. 问: 至少钓多少次鱼, 才能使总重量不少200kg的概率为0.95?

解 设此人共钓n次, 各次钓到的鱼的重量为随机变量X_i, 则

$E(X_i)=2, D(X_i)=2.25.$

令 $Z=\sum_{i=1}^{n} X_i$, 则$E(Z)=2n, D(Z)=2.25n.$

根据**林德贝格-列维中心极限定理,** Z近似服从 $N(2n, 2.25n)$.

目 上 下 回 停 续

则有

$$P\{Z \geq 200\} = P\left\{\frac{Z-2n}{\sqrt{2.25n}} \geq \frac{200-2n}{\sqrt{2.25n}}\right\}$$

$$= \Phi\left(-\frac{200-2n}{\sqrt{2.25n}}\right) = 0.95.$$

查表得 $\dfrac{200-2n}{1.5\sqrt{n}} = -1.645.$ 即n满足方程

$$n - 1.23375\sqrt{n} - 100 = 0$$

解方程, 得n=113.12. 因此, 取n=114即可.

目 上 下 回 停 续

例3-1 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. 且

$$E(X) = np = 170,$$

$$D(X) = np(1-p) = 170 \times 0.983$$

由**棣莫佛–拉普拉斯定理**知

保险公司亏本的概率为

$$
\begin{aligned}
&P\{10000X > 10000\times 200\}\\
&= P\{X > 200\}\\
&= P\left\{\frac{X-np}{\sqrt{np(1-p)}} > \frac{200-np}{\sqrt{np(1-p)}}\right\}\\
&= P\left\{\frac{X-170}{\sqrt{170\times 0.983}} > 2.321\right\}\\
&\approx 1-\Phi(2.321)\approx 0.01
\end{aligned}
$$

例3-2 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 º的概率为1/3, 若船舶遭受了90000次波浪冲击, 问其中有29500～30500次纵摇角大于 3 º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90000次波浪冲击中纵摇角大于 3 º的次数为X, 则X是一个随机变量, 且 $X \sim B(90000, 1/3)$. 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**棣莫佛-拉普拉斯定理**

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

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

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

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

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

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

$$=0.9995.$$

目 上 下 回 停 续

例3-3 某单位有1000部内线电话, 每部电话打外线的概率为0.05, 问需要装多少外线, 才能保证每部电话打外线时, 即时接通的概率不小于0.95?

解 令X表示同时要外线的电话机数, 则 $X \sim B(1000, 0.05)$, 且 $np=50$, $np(1-p)=47.5$.

根据**棣莫佛–拉普拉斯定理**, X近似服$N(50,47.5)$.

假定安装 k 条外线, 可使

$$P\{X \le k\} \ge 0.95$$

则有

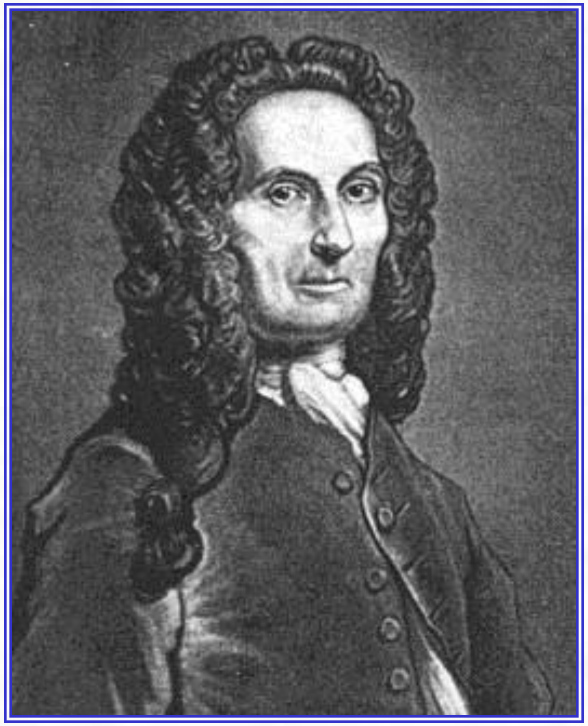
$$P\left\{\frac{X-50}{\sqrt{47.5}} \le \frac{k-50}{\sqrt{47.5}}\right\} \approx \Phi\left(\frac{k-50}{\sqrt{47.5}}\right) \ge 0.95.$$

查表得 $\Phi(1.645) = 0.95$. 由单调性, 应有

$$\frac{k-50}{\sqrt{47.5}} \ge 1.645,$$

解得 $k \ge 61.3$. 因此, 安装62条外线即可.

棣莫佛(Abraham de Moivre)



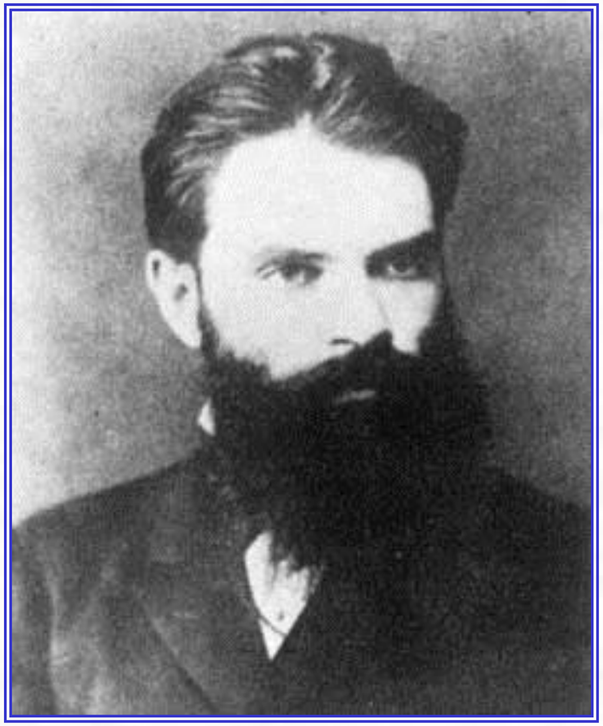
1667-1754

法国数学家.

发现了棣莫佛公式, 将复数与三角学联系起来.

主要的贡献是在一般分布与概率论上, 包括斯特林公式以及棣莫佛-拉普拉斯定理.

李雅普诺夫(Aleksandr Mikhailovich Lyapunov)



1857-1918

俄国数学家、力学家, 是切比谢夫创立的彼得堡学派的杰出代表.

在概率论方面, 创立了的特征函数方法, 实现了概率论极限理论在研究方法上的突破.

是常微分方程运动稳定性理论的创始人.

拉普拉斯(Pierre-Simon Laplace)

1749-1827

法国著名的天文学家和数学家,天体力学的集大成者.

因著名杰作《天体力学》被誉为是法国的牛顿.首次提出"天体力学"这一学科名称.

是现在广泛应用于各个领域的拉普拉斯变换和拉普拉斯方程的发现者.

定理4.9 李雅普诺夫(Liapunov)定理

设随机变量$X_1, X_2,\ldots, X_n$相互独立，它们具有数学期望与方差

$$E(X_i)=\mu_i,\quad D(X_i)=\sigma_i^2\quad (i=1,2,\cdots,n)$$

记 $B_n^2=\sum_{i=1}^{n}\sigma_i^2$，若存在正数$\delta$，使得当$n\to\infty$时

$$\frac{1}{B_n^{2+\delta}}\sum_{i=1}^{n}E\left\{\left|X_i-\mu_i\right|^{2+\delta}\right\}\to 0$$

目 上 下 回 停 续

则随机变量

$$Y_n^* = \frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu_i}{B_n}$$

的分布函数$F_n(x)$ 对于任意 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left\{Y_n^* \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t$$

注 1°　定理4.9是独立不同分布情形的中心极限定理, 该定理表明: 当n充分大时, 有

$$Y_n^* \sim AN(0,1)$$

而

$$\sum_{i=1}^{n} X_i \sim AN\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$$

2°　由**定理4.8**及**定理4.9**可以看出, 正态随机变量的普遍性及其在概率论中所占有的重要地位.

例2 一份考卷由99个题目组成, 并按由易到难顺序排列. 某学生答对1题的概率是 0.99; 答对第2题的概率是0.98; 一般地, 他答对第 i 题的概率是$1-\frac{i}{100}$ $(i=1, 2,\ldots, 99)$, 假如该学生回答各问题是相互独立的, 并且要正确回答其中60个问题以上(包括60)才算通过考试. 试计算该学生通过考试的概率是多少?

解 设

$$X_i=\begin{cases}1, & \text{学生答对第}i\text{题}\\ 0, & \text{学生答错第}i\text{题}\end{cases} \quad (i=1,2,\cdots,99)$$

目 上 下 回 停 续

于是 X_i 是两点分布:

$$P\{X_i=1\}=p_i, \quad P\{X_i=0\}=1-p_i$$

为了使其成为随机变量序列, 我们规定从 X_{100}开始都与X_{99}同分布, 且相互独立, 于是

$$B_n^2=\sum_{i=1}^{n} D(X_i)=\sum_{i=1}^{n} p_i(1-p_i)\to\infty \quad (n\to\infty)$$

另一方面, 因为

$$E\left(\left|X_i-p_i\right|^3\right)=p_i(1-p_i)^3+p_i^3(1-p_i)$$

$$=p_i(1-p_i)\left[p_i^2+(1-p_i)^2\right]\le p_i(1-p_i)<\infty$$

目 上 下 回 停 续

计算得

$$\sum_{i=1}^{99} E(X_i) = \sum_{i=1}^{99}\left(1-\frac{i}{100}\right) = 99-\frac{1}{100}\times\frac{99\times100}{2} = 49.5$$

$$B_{99}^2 = \sum_{i=1}^{99} D(X_i) = \sum_{i=1}^{99}\left(1-\frac{i}{100}\right)\left(\frac{i}{100}\right)$$

$$= 49.5-\frac{1}{100^2}\sum_{i=1}^{99} i^2$$

$$= 49.5-\frac{1}{100^2}\times\frac{99\times100\times199}{6} = 16.665$$

于是

$$\frac{1}{B_n^3}\sum_{i=1}^{n}E\left(\left|X_i-p_i\right|^3\right)\leq\frac{1}{\left[\sum_{i=1}^{n}p_i(1-p_i)\right]^{\frac{1}{2}}}\to 0$$

$$(n\to\infty)$$

即独立随机变量序列满足**李雅普诺夫定理**的条件.

因此随机变量 $Y_n^* = \dfrac{\sum\limits_{i=1}^{n}X_i-\sum\limits_{i=1}^{n}E(X_i)}{\sqrt{D\left(\sum\limits_{i=1}^{n}X_i\right)}}$

近似服从标准正态分布 $N(0, 1)$.

目 上 下 回 停 续