第三节　中心极限定理

一、依分布收敛

二、中心极限定理

三、小结

5.3.1 依分布收敛

定义　设 $\boldsymbol{X}, \boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n, \cdots$ 的分布函数依次为

$$\boldsymbol{F}(\boldsymbol{x}), \boldsymbol{F}_1(\boldsymbol{x}), \boldsymbol{F}_2(\boldsymbol{x}), \cdots, \boldsymbol{F}_n(\boldsymbol{x}), \cdots$$

如果对于$F(x)$ 的每个连续点 x，都有

$$\lim_{n\to\infty} \boldsymbol{F}_n(\boldsymbol{x}) = \boldsymbol{F}(\boldsymbol{x})$$

则称随机变量序列 $\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_n, \cdots$ 依分布收敛于 X，记为

$$X_n \xrightarrow{L} X \quad (n \to \infty)$$

注　依分布收敛概念给出了随机变量序列对应的分布函数序列的极限与某一特定随机变量分布函数之间的关系.

5.3.2 中心极限定理

1. 为何正态分布在概率论中地位非凡？

2. 大样本统计推断的理论基础是什么？

独立同分布（Lévy-Lindberg）中心极限定理

棣莫弗-拉普拉斯中心极限定理

李雅普诺夫中心极限定理

定理一（独立同分布的中心极限定理）

莱维-林德伯格（Lévy-Lindberg）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

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

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

独立同分布中心极限定理可叙述成:

设 $X_1, X_2, \cdots, X_n, \cdots$相互独立, 服从同一分布, 存在期望$E(X_k)=\mu$ 和方差 $D(X_k)=\sigma^2>0\ (k=1,2,\cdots)$

则 $Y_n=\dfrac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$ 依分布收敛于标准正态分布 $N(0,1)$,

注

n 足够大时, Y_n的分布函数近似于标准正态随机变量的分布函数, 且 $\sum\limits_{k=1}^{n}X_k=\sqrt{n}\sigma Y_n+n\mu$ 近似服从 $N(n\mu,n\sigma^2)$

例5.3.1 一射击运动员在一次射击中所得环数 X 的概率分布如下：

X	10	9	8	7	6
P	0.5	0.3	0.1	0.05	0.05

问在100次独立射击中所得总环数介于900与930环之间的概率是多少？

解 设 X_k 表示第 k 次射击所得环数，则 $X_1, X_2, \cdots, X_{100}$ 相互独立，都与 X 同分布.

$$\mu = E(X_k) = E(X) = 9.15,\ \sigma^2 = D(X_k) = D(X) = 1.23$$

100 次射击所得总环数为 $\sum_{k=1}^{100} X_k$，据独立同分布的中心极限定理

$$\frac{\sum_{k=1}^{100} X_k - 100\times 9.15}{\sqrt{100\times 1.23}}$$ 近似服从 $N(0,1)$，因此所求概率为

$$P\{900 \le \sum_{k=1}^{100} X_k \le 930\}$$

$$= \boldsymbol{P}\left\{\frac{900-100\times 9.15}{\sqrt{100\times 1.23}} \le \frac{\sum_{k=1}^{100} \boldsymbol{X}_k - 100\times 9.15}{\sqrt{100\times 1.23}} \le \frac{930-100\times 9.15}{\sqrt{100\times 1.23}}\right\}$$

$$\approx \boldsymbol{P}\left\{-1.35 \le \frac{\sum_{k=1}^{100} \boldsymbol{X}_k - 100\times 9.15}{\sqrt{100\times 1.23}} \le 1.35\right\}$$

$$= \Phi(1.35) - \Phi(-1.35)$$

$$= 2\Phi(1.35) - 1$$

$$= 2\times 0.9115 - 1 = 0.8230 .$$

例5.3.1续 一射击运动员在一次射击中所得环数 X 的概率分布如下：

X	10	9	8	7	6
P	0.5	0.3	0.1	0.05	0.05

(2) 利用切比雪夫不等式估计在100次独立射击中所得总环数介于900与930环之间的概率。

解 设 X_k 表示第 k 次射击所得环数, 则 $X_1, X_2, \cdots, X_{100}$ 相互独立, 都与 X 同分布.

$$\mu = E(X_k) = E(X) = 9.15,\ \sigma^2 = D(X_k) = D(X) = 1.23$$

100 次射击所得总环数为 $\sum_{k=1}^{100} X_k \triangleq Y$, 由切比雪夫不等式,

$$P\{900 < Y < 930\} = P\{-15 < Y - 915 < 15\} = P\{|Y - 915| < 15\}$$

$$\geq 1 - \frac{123}{15^2} = 0.45.$$

例5.3.2 一生产线生产的产品成箱包装，每箱重量随机，设每箱平均重50kg,标准差为5kg. 若用最大载重量为5t的汽车承运，试利用中心极限定理说明：每辆车最多可以装多少箱，才能保障不超重的概率大于0.9772.

解 设n是所求箱数，X_i是装运的第i箱重量($i=1,2,\ldots,n$).

$X_1, X_2, \cdots, X_n$独立同分布且$E(X_i)=50, \sqrt{D(X_i)}=5, i=1,2\cdots,n.$

由中心极限定理，$\sum_{i=1}^{n} X_i$近似服从$N(50n, 25n)$.

$$P\{\sum_{i=1}^{n} X_i \le 5000\} = P\{\frac{\sum_{i=1}^{n} X_i - 50n}{5\sqrt{n}} \le \frac{5000-50n}{5\sqrt{n}}\} = \Phi(\frac{1000-10n}{\sqrt{n}}) > 0.9772,$$

由于$\Phi(x)$单调增加，$\Phi(2)=0.9772$,

所以$\frac{1000-10n}{\sqrt{n}} > 2$，解得$n < 98.0199$,

所以最多可以装98箱。

定理二(棣莫佛—拉普拉斯定理)

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

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

证明 根据第四章例题可知 $Y_n = \sum_{k=1}^{n} X_k$,

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

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

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

根据独立同分布的中心极限定理得

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

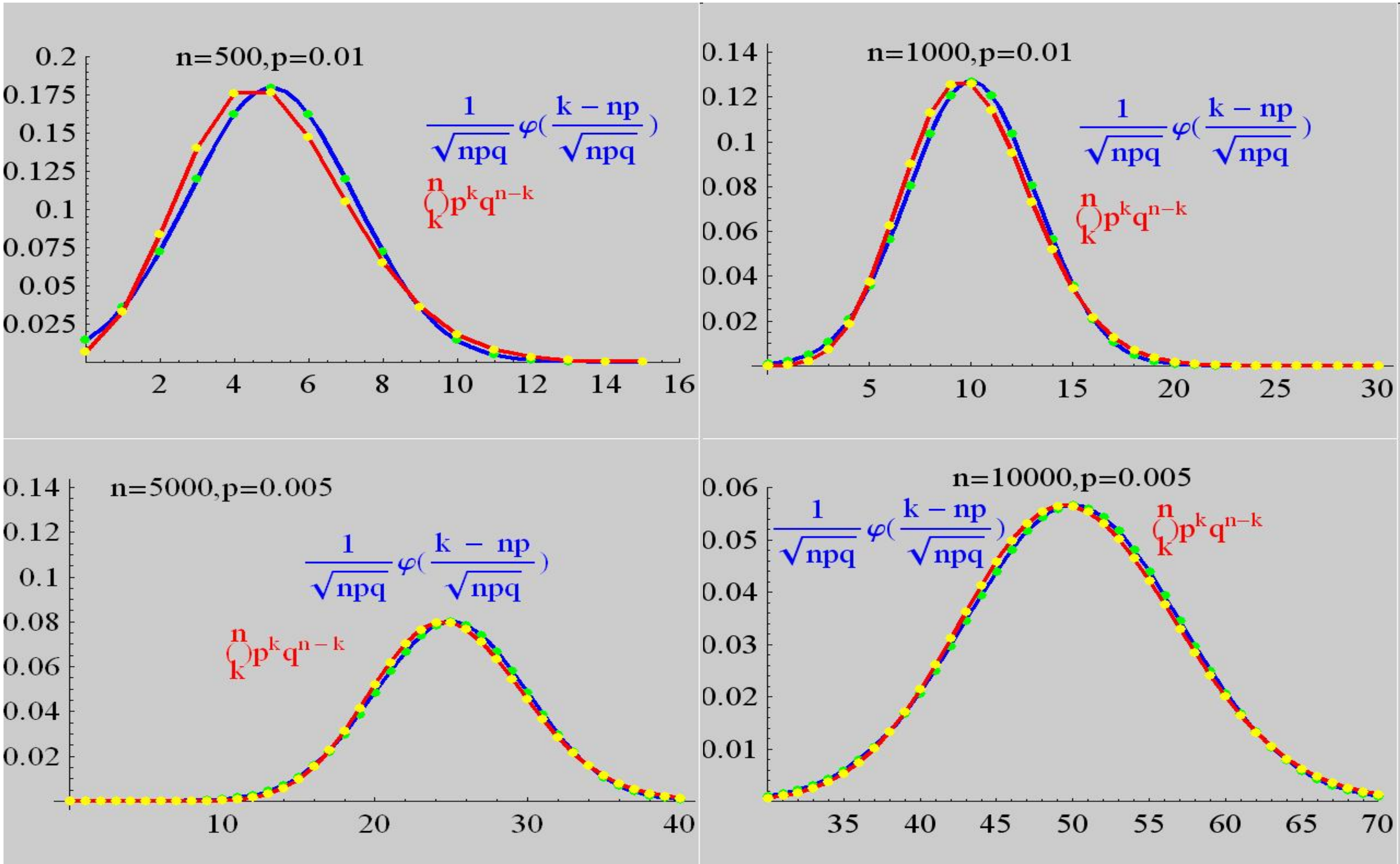
$$= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi(x).$$

即 $\dfrac{Y_n - np}{\sqrt{np(1-p)}}$ 依分布收敛于标准正态分布 $N(0, 1)$,

相应地Y_n依分布收敛于$N(np, np(1-p))$.

注：棣莫佛一拉普拉斯定理是独立同分布中心极限定理的特例。

下面的图形表明:正态分布是二项分布的极限分布.



只要 n 充分大, 可用正态分布近似计算二项分布的概率.

棣莫佛-拉普拉斯定理表明，当n很大，$0<p<1$是一个定值时，或者说$np(1-p)$也不太大时，二项分布$B(n,p)$近似服从正态分布$N(np,np(1-p))$.

泊松定理表明，当n很大，p很小($p<0.1$)时，二项分布$B(n,p)$近似服从参数为np的泊松分布。

例5.3.3 某工厂有200台同类型机器，由于工艺等原因，每台机器实际工作时间只占全部工作时间的75%，假定每台机器工作相互独立. 求任一时刻有144至160台机器正在工作的概率。

解 以X表示任一时刻正在工作的机器台数，

则$X \sim B(200, 0.75)$. 由棣莫佛一拉普拉斯定理有

$$P\{144 \le X \le 160\} = P\left\{\frac{144-200\times0.75}{\sqrt{200\times0.75\times0.25}} \le \frac{X-200\times0.75}{\sqrt{200\times0.75\times0.25}} \le \frac{160-200\times0.75}{\sqrt{200\times0.75\times0.25}}\right\}$$

$$= P\left\{\frac{144-150}{\sqrt{37.5}} \le \frac{X-150}{\sqrt{37.5}} \le \frac{160-150}{\sqrt{37.5}}\right\}$$

$$= P\left\{\frac{X-150}{\sqrt{37.5}} \le 1.63\right\} - P\left\{\frac{X-150}{\sqrt{37.5}} \le -0.98\right\}$$

$$\approx \Phi(1.63) - \Phi(-0.98) = \Phi(1.63) - [1-\Phi(0.98)]$$

$$= 0.9484 - (1-0.8365) = 0.7849.$$

定理三(李雅普诺夫定理*)

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:

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

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

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

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

说明：

1. 在李雅普诺夫定理的条件下，当n很大时，随机变量

$$Z_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}\mu_k}{B_n}$$近似服从标准正态分布$N(0,1)$.

2. 当n很大时，$\sum_{k=1}^{n}X_k=B_nZ_n+\sum_{k=1}^{n}\mu_k$近似服从正态分布$N(\sum_{k=1}^{n}\mu_k,B_n^2)$.

3. 当n很大时，无论各随机变量$X_k\ (k=1,2,\cdots)$服从什么分布，

$\sum_{k=1}^{n}X_k$就近似服从正态分布.

4. 许多实际问题中，所考察随机变量可视作多个独立随机变量之和，

如城市耗电量、物理实验测量误差等，它们都近似服从正态分布.

三、小结

三个中心极限定理

- 独立同分布的中心极限定理
- 棣莫佛－拉普拉斯定理
- 李雅普诺夫定理

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.

大数定律和中心极限定理的区别和联系：

区别：大数定律研究随机变量序列之和依概率收敛的极限问题，而中心极限定理是研究随机变量序列之和依分布收敛的极限定理。

联系：都是研究随机变量和的极限行为。