概率论与数理统计

第二节　　中心极限定理

- 中心极限定理
- 例题

中心极限定理的客观背景

在实际问题中许多随机变量是由相互独立随机因素的综合（或和)影响所形成的.

例如：炮弹射击的落点与目标的偏差，就受着许多随机因素（如瞄准，空气阻力，炮弹或炮身结构等）综合影响的. 每个随机因素的对弹着点（随机变量和）所起的作用都是很小的. 那么弹着点服从怎样分布哪 ？

自从高斯指出测量误差服从正态分布之后,人们发现,正态分布在自然界中极为常见.

如果一个随机变量是由大量相互独立的随机因素的综合影响所造成,而每一个别因素对这种综合影响中所起的作用不大. 则这种随机变量一般都服从或近似服从正态分布.

“多因素,小影响,综合为正态。”

一、中心极限定理

定理1（独立同分布下的中心极限定理）

设随机变量$X_1, X_2, \cdots X_n, \cdots$独立同分布，且具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2$ $(k=1,2,\cdots)$，则随机变量之和$\sum_{k=1}^{n} X_k$的标准化变量

$\dfrac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma}$的分布函数$F_n(x)$对于任意$x$满足

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

注 1、定理表明，独立同分布的随机变量之和$\sum_{k=1}^{n} X_k$，

当n充分大时，随机变量之和的标准化变量

$$\frac{\sum_{k=1}^{n} X_k - n\mu}{\sqrt{n}\sigma} \overset{\text{近似地}}{\sim} N(0,1); \sum_{k=1}^{n} X_k \overset{\text{近似地}}{\sim} N(n\mu, n\sigma^2).$$

2、独立同分布中心极限定理的另一种形式可写为

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似地}}{\sim} N(0,1); \text{或} \quad \bar{X} \overset{\text{近似地}}{\sim} N(\mu, \sigma^2/n) \text{其中} \bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

3、虽然在一般情况下,我们很难求出$\sum_{k=1}^{n} X_k$的分布的确切形式,但当n很大时,可以求出近似分布.

定理5.6（棣莫佛—拉普拉斯定理）

在n重贝努利试验中，成功的次数为Y_n，而在每次试验中，试验成功的概率为p，$q=1-p$ 则对一切x有

$$\lim_{n\to\infty} P\{\frac{Y_n-np}{\sqrt{npq}}\le x\}=\Phi(x)$$

$$\frac{Y_n-np}{\sqrt{npq}} \overset{\text{近似地}}{\sim} N(0,1);$$

二、中心极限定理例题

例1 某人要测量A,B两地之间的距离,限于测量工具,将其分成1200段进行测量,设每段测量误差(单位km)相互独立,且均服从$(-0.5,0.5)$上的均匀分布。试求距离测量误差X的绝对值不超过$20km$ 的概率

解 (1)以$X_k(k=1,2\cdots,1200)$记第k段距离测量误差,

$X_k \sim U(-0.5,0.5)$, 易知$E(X_k)=0$, $D(X_k)=\dfrac{1}{12}$,

$k=1,2,\cdots 1200.$ 而$X=\sum\limits_{k=1}^{1200} X_k$.

由中心极限定理可知

$$P\{|X|\le 20\}=P\left\{\left|\sum_{k=1}^{1200}X_k\right|\le 20\right\}$$

$$\approx P\left\{\frac{-20}{10}\le\frac{\sum_{k=1}^{1200}X_k-1200\times 0}{\sqrt{1200\times\frac{1}{12}}}\le\frac{20}{10}\right\}$$

$$=\Phi(2)-\Phi(-2)\approx 0.9$$

例2 一个复杂系统由100个相互独立的元件组成。在系统运行期间每个元件损坏的概率为0.10 ,又知道为使系统正常运行,至少必须有85个元件作,

(1)求系统的可靠性度（即正常工作的概率）

(1)设X 表示系统正常运行时完好的元件个数

$$X_k=\begin{cases}1 & \text{第}k\text{个元件没损坏}\\ 0 & \text{第}k\text{个元件损坏}\end{cases}\qquad \text{而}X=\sum_{k=1}^{1200}X_k.$$

易知$E(X_k)=0.9, D(X_k)=0.09, \quad k=1,2,\cdots 100.$

由中心极限定理可知

$$P\{X>85\}=P\left\{\sum_{k=1}^{100}X_k>85\right\}$$

$$\approx P\left\{\frac{\sum_{k=1}^{100}X_k-100\times 0.9}{\sqrt{100\times 0.09}}>\frac{85-90}{3}\right\}$$

$$=1-\Phi(-\frac{5}{3})\approx 0.952$$

例2 (2)假设上述系统由n个相互独立的元件组成，而且要求至少有80%的元件工作才能使整个系统正常运行,问n至少多大时候,才能保证系统的可靠度为0.95

$$P\{X \geq 0.8n\} = P\left\{\sum_{k=1}^{n} X_k \geq 0.8n\right\}$$

$$\approx P\left\{\frac{\sum_{k=1}^{n} X_k - n \times 0.9}{\sqrt{n \times 0.09}} > \frac{0.8n - 0.9n}{0.3\sqrt{n}}\right\} = \Phi\left(\frac{\sqrt{n}}{3}\right) \geq 0.95$$

$n=25$

例3 设随机变量$X_1, X_2, \cdots X_n, \cdots$独立同分布，且$X_k(k=1,2,\cdots)$服从参数为$\beta$的指数分布,则() **(A)**

(A) $\lim\limits_{n\to\infty} P\left\{\dfrac{\beta\sum\limits_{i=1}^{n} X_i - n}{\sqrt{n}} \le x\right\} = \Phi(x)$ (B) $\lim\limits_{n\to\infty} P\left\{\dfrac{\sum\limits_{i=1}^{n} X_i - n}{\sqrt{n}} \le x\right\} = \Phi(x)$

(C) $\lim\limits_{n\to\infty} P\left\{\dfrac{\sum\limits_{i=1}^{n} X_i - \beta}{\sqrt{n\beta}} \le x\right\} = \Phi(x)$ (D) $\lim\limits_{n\to\infty} P\left\{\dfrac{\sum\limits_{i=1}^{n} X_i - \beta}{n\beta} \le x\right\} = \Phi(x)$

易知$E(X_i) = \dfrac{1}{\beta}$, $D(X_i) = \dfrac{1}{\beta^2}$