西安電子科技大學
XIDIAN UNIVERSITY

第5章 大数定律及中心极限定理



5.3 中心极限定理

在许多实际问题中，我们会经常遇到这样的随机变量，它是由大量的相互独立的随机因素的综合影响而形成的，而其中每一个个别因素在总的影响中所起的作用是微小的，这种随机变量往往近似地服从正态分布，这就是中心极限定理的实际背景。

1、独立同分布中心极限定理

定理 1 设 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立同分布的随机变量序列，且具有如下数学期望和方差：

$$EX_i = \mu, DX_i = \sigma^2 (i = 1, 2, \cdots)$$

则 $\forall x \in R$,随机变量 $Y_n = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{D(\sum_{i=1}^n X_i)}} = \frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma}$ 的分布函数 $F_n(x)$ 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

定理1在实际中有着广泛的应用。当n充分大时，$\sum_{i=1}^n X_i$ 近似服从以它的均值为均值，它的方差为方差的正态分布，即 $\sum_{i=1}^n X_i$ 近似地服从正态分布$N(n\mu, n\sigma^2)$。因此对于任意的实数x及$a<b$, 有

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

$$P\left(a < \sum_{i=1}^{n} X_i \le b\right) \approx \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right)$$

2、 *Lyapunov*中心极限定理

定理 2 设$X_1, X_2, \dots, X_n, \dots$是相互独立的随机变量序列，且具有如下的数学期望和方差：

$$EX_i = \mu_i, DX_i = \sigma_i^2, i = 1, 2, \dots$$

记

$$B_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

若存在 $\delta > 0$，使得当 $n \to \infty$ 时

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

则$\forall x \in R$, 随机变量 $Y_n = \frac{\sum_{i=1}^n X_i - E(\sum_{i=1}^n X_i)}{\sqrt{D(\sum_{i=1}^n X_i)}} = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{B_n}$ 的分布函数$F_n(x)$

满足 $\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\left(\frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{B_n} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \Phi(x)$

2、 *Lyapunov*中心极限定理

定理 2 表明：无论$X_1, X_2, \dots, X_n, \dots$服从什么分布，只要满足定理2的条件，那么当n充分大时，$\sum_{i=1}^{n} X_i$近似地服从以它的均值为均值, 以它的方差为方差的正态分布。

即：$\sum_{i=1}^{n} X_i$ 近似地服从正态分布 $N\left(\sum_{i=1}^{n}\mu_i, \sum_{i=1}^{n}\sigma_i^2\right)$。

由定理1容易推出历史上著名的 *DeMoivre Laplace*（棣莫弗 拉普拉斯）中心极限定理，它在二项分布随机变量形成事件概率的近似计算中有着重要的作用。

定理 3（*DeMoivre Laplace*中心极限定理）设n_A表示n重*Bernoulli*试验中事件A发生的次数，p是事件A在一次试验中发生的概率，即$P(A)=p$，则$\forall x\in R$，

随机变量$Y_n=\frac{n_A-np}{\sqrt{np(1-p)}}$的分布函数$F_n(x)$满足

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

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

该定理表明：若$X\sim B(n,p)$，当n充分大时，则 X近似地服从以它的均值为均值，它的方差为方差的正态分布，即X近似地服从正态分布$N(np,np(1-p))$。因此，对于任意的实数x及$a<b$，有

$$P(X\leq x)\approx\Phi\left(\frac{x-np}{\sqrt{np(1-p)}}\right)$$

$$P(a<X\leq b)\approx\Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right)$$

例1 调查吸烟率p，问要调查多少人才能保证吸烟频率与p的差不超过0.005的概率不低于95%？

现用中心极限定理求解：

$$P\left\{\left|\frac{n_A}{n}-p\right|\le 0.005\right\}\ge 0.95 \qquad n_A\sim B(n,p)$$

$$P\left\{\left|\frac{n_A-np}{\sqrt{np(1-p)}}\right|\le\frac{0.005\sqrt{n}}{\sqrt{p(1-p)}}\right\}\ge 0.95$$

左式 $\approx\Phi\left(\dfrac{0.005\sqrt{n}}{\sqrt{p(1-p)}}\right)-\Phi\left(\dfrac{-0.005\sqrt{n}}{\sqrt{p(1-p)}}\right)$

估计$(1-p)p$的范围后，请自行查表完成计算。

例2 有100道考试题，每题$ABCD$四选一，选对得1分，选错不得分。

（1）现随机乱选，求得分大于20分和得分大于40分的概率。

（2）若经过模拟考试，选对的可能性提高到85%，求考试得90分以上的概率。

解

（1）设 $X_i=\begin{cases}1 & 第i题得分\\ 0 & 其他\end{cases}$ $\quad X=X_1+X_2+\cdots+X_n\ ,\ n=100$

$X\sim B(100,p)$

$$E(X)=np=100\times\frac{1}{4}=25\ ,\ D(X)=np(1-p)=100\times\frac{1}{4}\times\frac{3}{4}=\frac{75}{4}$$

$$P(X\geq 20)=P\left(\frac{X-25}{\sqrt{75/4}}\geq\frac{20-25}{\sqrt{75/4}}\right)\approx 1-\Phi\left(\frac{-5}{\sqrt{75/4}}\right)=\Phi\left(\frac{2}{\sqrt{3}}\right)\approx 87.5\%$$

$$P(X\geq 40)\approx 1-\Phi(2\sqrt{3})\approx 0\%$$

（2）$p=0.85,np=85,np(1-p)=85\times 0.15$

$$P(X\geq 90)=P\left(\frac{X-85}{\sqrt{85\times 0.15}}\geq\frac{90-85}{\sqrt{85\times 0.15}}\right)\approx 1-\Phi\left(\frac{5}{\sqrt{85\times 0.15}}\right)\approx 0.08$$

超过90分有8%的可能性。

$$P(X\geq 91)\approx 1-\Phi\left(\frac{6}{\sqrt{85\times 0.15}}\right)\approx 0.046<5\%$$

例3

（1）$X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从（−1, 1）上的均匀分布 $\lim\limits_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum\limits_{i=1}^{n} X_i \le 1\right) = ?$

（2） $X_1, X_2, \cdots, X_{100}$ 相互独立，服从参数为4的泊松分布 $P\left(\overline{X} \le 4.392\right) = ?$

解

（1） $E(X_i) = 0, D(X_i) = \frac{1}{3}$ $\qquad E(\sum\limits_{n} X_i) = 0, D(\sum\limits_{n} X_i) = \frac{n}{3}$

$$\lim_{n\to\infty} P\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i \le 1\right) = \lim_{n\to\infty} P\left(\frac{\sum\limits_{n} X_i}{\sqrt{\frac{n}{3}}} \le \sqrt{3}\right) = \Phi\left(\sqrt{3}\right)$$

（2） $E(X_i) = 4, D(X_i) = 4$ $\qquad E(\overline{X}) = 0, D(\overline{X}) = \frac{4}{100} = 0.04$

$$P\left(\overline{X} \le 4.392\right) = P\left(\frac{\overline{X} - 4}{0.2} \le \frac{4.392 - 4}{0.2}\right) \approx \Phi\left(\frac{4.392 - 4}{0.2}\right) = \Phi(1.96)$$

例4 某车间有200台车床，由于各种原因每台车床只有60%的时间在开动，每台车床开动期间耗电量为1单位，问至少供给此车间多少电量才能以不少于99.9%的概率保证此车间不因供电不足而影响生产。

解 设不影响生产需要开动的车床数为n，X表示200台车床中开动的车床数，则$X\sim B(200,0.6)$，从而

$$EX=200\times 0.6=120, DX=200\times 0.6\times 0.4=48$$

由中心极限定理知，X近似地服从正态分布N (120，48)，所以n取决于如下条件：

$$P(X\le n)\approx \Phi\left(\frac{n-120}{\sqrt{48}}\ge 0.999\right)$$

查表得$\frac{n-120}{\sqrt{48}}\ge 3.01$，即$n$≥141，故至少需供给此车间141单位的电量才能以不少于99.9%的概率保证此车间不因供电不足而影响生产。