§5.2 中心极限定理

一. 中心极限定理的定义与意义

定义5.2.1 设随机变量$X, X_1, X_2, \ldots$的分布函数分别为$F(x), F_1(x), F_2(x), \ldots$, 若极限式

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

在$F(x)$的每一个连续点上都成立，称随机变量序列$\{X_k\}, k = 1,2,\ldots$依分布收敛于X.

记为 $X_n \xrightarrow{L} X$

定义5.2.2（中心极限定理）设随机变量序列$\{X_k\}$，$k=1,2,\ldots$相互独立，有有限数学期望和方差.若随机变量序列

标准化

$$Y_n=\frac{\sum_{k=1}^{n}X_k-\sum_{k=1}^{n}E(X_k)}{\sqrt{\sum_{k=1}^{n}D(X_k)}}$$

对$y\in R$一致地有

$$\lim_{n\to\infty} P\{Y_n < y\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{1}{2}t^2}dt = \Phi(y)$$

称随机变量序列 $\{X_k\}$服从**中心极限定理.**

注1 随机变量序列 $\{X_k\}$服从中心极限定理，指其前n项和 $\sum_{k=1}^{n} X_k$的标准化随机变量依分布收敛于标准正态分布随机变量X；

注2 解释了现实中哪些随机变量可看服从正态分布；

若随机变量序列$\{X_k\}$，$k = 1,2,\ldots$服从中心极限定理，有

$$\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} \xrightarrow{L} X \sim N(0,1) \quad as\ n \to \infty$$

故当n 足够大时，可以认为

$$\frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} \sim N(0,1)$$

近似成立，或

$$\sum_{k=1}^{n} X_k \sim N\left(\sum_{k=1}^{n} E(X_k), \sum_{k=1}^{n} D(X_k)\right)$$

近似成立.

许多相互独立的微小因素X_k的叠加总和.

注3 给出了概率的近似计算公式.

若随机变量序列$\{X_k\}$，$k=1,2,\ldots$服从中心极限定理，则有

$$P\left\{x_1 \le \frac{\sum_{k=1}^{n} X_k - \sum_{k=1}^{n} E(X_k)}{\sqrt{\sum_{k=1}^{n} D(X_k)}} < x_2\right\} \approx \Phi(x_2) - \Phi(x_1)$$

二. 中心极限定理

定理5.2.1（林德伯格—列维定理、独立同分布中心极限定理)

设$\{X_k\}, k=1,2\ldots$为相互独立, 具有相同分布的随机变量序列, 且$E(X_k)=\mu, D(X_k)=\sigma^2$, 则$\{X_k\}$满足中心极限定理，即 有

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

重复试验次数估计

装车问题

报亭售报问题

定理5.2.2（棣莫佛-拉普拉斯中心极限定理）

设随机变量序列$\{Y_n\}$，$Y_n \sim B(n, p)$，$n=1,2\ldots$，对于任意的实数x，有

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

证明 对于任意正整数n，随机变量Y_n可表示为

$$Y_n = X_1 + X_2 + \ldots + X_n$$

其中$X_i \sim B(1, p)$，相互独立，并且

$$E(X_i) = p \quad, D(X_i) = p(1-p)$$

相互独立同分布的随机变量序列$\{X_i\}, i = 1,2,\ldots$满足中心极限定理. 即有

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

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

结论成立.

若$X \sim B(n, p)$，**对于足够大的**n，**有**

$$P\{m_1 < X \le m_2\}$$

标准化

$$= P\left\{ \frac{m_1 - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \le \frac{m_2 - np}{\sqrt{np(1-p)}} \right\}$$

$$\approx \Phi\left(\frac{m_2-np}{\sqrt{np(1-p)}}\right)-\Phi\left(\frac{m_1-np}{\sqrt{np(1-p)}}\right)$$

产品抽检件数

中心极限定理应用实例

例5.2.1 将一枚均匀硬币连续抛 n 次，试用中心极定理来估计 n ，使下式成立.

$$P\{|f_n(A)-P(A)|<0.01\}\geq 0.99$$

其中 A ={ 出现正面 }

解 有$P(A)$=1/2，令

$$X_i=\begin{cases}1, & 第i次出现正面;\\ 0, & 否则,\end{cases} \qquad (i=1,2,\cdots n)$$

则随机变量序列$\{X_i\}$，$i=1,2,\ldots$相互独立且同分布. 而且有

$$E(X_i)=\frac{1}{2},\quad D(X_i)=\frac{1}{4},\quad i=1,2,\cdots$$

所以随机变量序列$\{X_i\}$，满足独立同分布中心极限定律.

对$f_n(A)=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$,由题意可得

$$0.99 \le P\{|f_n(A)-P(A)|<0.01\}$$

$$=P\left\{\frac{1}{2}-0.01<\frac{1}{n}\sum_{i=1}^{n}X_i<\frac{1}{2}+0.01\right\}$$

$$= P\left\{\frac{n}{2} - 0.01n < \sum_{i=1}^{n} X_i < \frac{n}{2} + 0.01n\right\}$$

$$= P\left\{-\frac{0.01n}{1/2\sqrt{n}} < \frac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{1/2\sqrt{n}} < \frac{0.01n}{1/2\sqrt{n}}\right\}$$

因为 $\dfrac{\sum_{i=1}^{n} X_i - \frac{n}{2}}{1/2\sqrt{n}}$ 近似服从$N(0,1)$分布，

所以$0.99 \le P\{-\dfrac{0.01n}{1/2\sqrt{n}} < \dfrac{\sum_{i=1}^{n} X_i - \dfrac{n}{2}}{1/2\sqrt{n}} < \dfrac{0.01n}{1/2\sqrt{n}}\}$

$$\approx 2\Phi(0.02\sqrt{n}) - 1$$

$$\Rightarrow \Phi(0.02\sqrt{n}) \ge 0.995$$

$$\Rightarrow 0.02\sqrt{n} \ge 2.58$$

解得 $n \ge 16{,}641$ (次) (250,000次)

例5.2.2 一生产线生产的产品成箱包装，每箱的重量是随机的.假设每箱平均重50千克，标准差为5千克. 若用最大载重量为5吨的汽车装运，试利用中心极限定理说明每辆车最多可以装多少箱，才能保障不超载的概率大于0.977.

解 设X_i，$i=1,2,\ldots,n$ 是装运的第i 箱重量(单位:千克), n是所求箱数.

n 箱的总重量为

$$T_n = X_1 + X_2 + \cdots + X_n$$

$$E(X_i) = 50, \quad \sqrt{D(X_i)} = 5,$$

$$E(T_n) = 50n, \quad \sqrt{D(T_n)} = 5\sqrt{n}, \quad (\text{单位}:\text{千克})$$

可将X_i，$i=1,2,\ldots,n$ **视为独立同分布的随机变量.**

由林德伯格—列维定理知，T_n **近似服从正态分布** $N(50n, 25n)$.

$$P\{T_n \le 5000\} = P\{\frac{T_n - 50n}{5\sqrt{n}} \le \frac{5000 - 50n}{5\sqrt{n}}\}$$

$$\approx \Phi(\frac{1000 - 10n}{\sqrt{n}}) > 0.977 = \Phi(2),$$

故 $$\frac{1000 - 10n}{\sqrt{n}} > 2,$$

解得 $n < 98.0199$, 即一辆车最多可以装98箱.

例5.2.3 路边有一个售报亭，每个过路人在报亭买报的概率是 1/3，求：正好售出 100 份报纸时的过路人数在 280 到 300 之间的概率。

解 设 X 是正好售出 100 份报纸时的过路人数，X_i 是售出第 $i-1$ 份报纸后到售出第 i 份报纸时的过路人数，则

$$X=\sum_{i=1}^{100}X_i$$

并且随机变量 $X_1, X_2, \cdots, X_{100}$ 独立同分布，具有分布律:

$$P\{X_i = k\} = \frac{1}{3}\left(\frac{2}{3}\right)^{k-1}, \qquad k = 1,2,\cdots\cdots$$

因

$$E(X_i) = \frac{1}{\frac{1}{3}} = 3, \qquad D(X_i) = \frac{\frac{2}{3}}{\left(\frac{1}{3}\right)^2} = 6$$

$$i = 1, 2, \cdots, 100;$$

根据林德伯格—列维定理，所求概率

$$P\{280 < \sum_{i=1}^{100} X_i < 300\}$$

$$\approx \Phi\left(\frac{300-100\times 3}{\sqrt{100\times 6}}\right) - \Phi\left(\frac{280-100\times 3}{\sqrt{100\times 6}}\right)$$

$$= \Phi(0) - \Phi(-0.8165)$$

$$= 0.5 - 1 + \Phi(0.8165)$$

$$= 0.293$$

例5.2.4 随机抽查验收产品，如果在一批产品中查出10个以上的次品，则拒绝接收. 问至少检查多少个产品，能保证次品率为10%的一批产品被拒收的概率不低于0.9

解 设检查的产品数为 n，查出的次品数为X，则$X \sim B(n,\ 0.1)$，按题意，有

$$P\{10 < X \le n\} \ge 0.9$$

由棣莫佛 - 拉普拉斯中心极限定理，有

$P\{\ 10 < X \leq n\ \}$

$$\approx \Phi\left(\frac{n-0.1n}{\sqrt{0.1\times 0.9n}}\right)-\Phi\left(\frac{10-0.1n}{\sqrt{0.1\times 0.9n}}\right)$$

$$=\Phi(3\sqrt{n})-\Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right)$$

于是

$$P\{10 < X \leq n\}\approx 1-\Phi\left(\frac{10-0.1n}{0.3\sqrt{n}}\right)$$

$$=\Phi\left(\frac{0.1n-10}{0.3\sqrt{n}}\right)$$

故

$$\Phi\left(\frac{0.1n-10}{0.3\sqrt{n}}\right) \geq 0.9$$

$$\Rightarrow \quad \frac{0.1n-10}{0.3\sqrt{n}} \geq 1.28$$

求解得 $n \geq 146.8$ 或 $n \leq -68.3$，

所以至少取 $n = 147$ 能够保证要求.

应用范例

在计算机模拟试验中，常利用12个相互独立同分布，都在（0，1）上服从均匀分布的随机变量$X_1, X_2, \ldots, X_{12}$之和的标准化

$$Y=\frac{[\sum_{i=1}^{12}X_i-6]}{\sqrt{12/12}}=\sum_{i=1}^{12}X_i-6$$

作为标准正态分布的随机变量.

根据林德伯格—列维定理，Y应近似服从标准正态分布.

事实上二者的概率密度几乎无区别.

例如X_1, X_2, X_3相互独立，都在(0，1)上均匀分布，则$S_3 = X_1+X_2+X_3$的概率密度为

$$f(x)=\begin{cases} \dfrac{x^2}{2}, & x\in(0,1); \\ -x^2+3x-\dfrac{3}{2}, & x\in[1,2); \\ \dfrac{(3-x)^2}{2}, & x\in[2,3); \\ 0, & \text{其他}. \end{cases}$$

中心极限定理

将S_3标准化： $\dfrac{S_3 - 3/2}{\sqrt{1/4}} = 2S_3 - 3$

其概率密度为： $g(x) = \dfrac{1}{2} f\left(\dfrac{x+3}{2}\right)$