第五章　大数定律和中心极限定理

极限定理研究大量随机变量的规律性.

1. 大数定律(Law of Large Numbers)：大量随机变量的平均结果的稳定性.
2. 中心极限定理(Central Limit Theorem)：大量随机变量的和的稳定性.

5.1　大数定律

定义. 设 $Y_1,...,Y_n$ 是一个随机变量序列，Y 是一个常数，若对于任给的 $\varepsilon>0$，有

$$\lim_{n\to\infty} P\{|Y_n-a|<\varepsilon\}=1,$$

则称随机变量序列 $\{Y_n\}$ 依概率收敛于 a，记作

$$Y_n \xrightarrow{P} a.$$

依概率收敛的序列有如下性质：

设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$, 又设 $g(x,y)$ 在点 (a,b) 连续, 则

$$g(X_n,Y_n) \xrightarrow{P} g(a,b).$$

定理. 设随机变量 X 有期望和方差，则对于任给的 $\varepsilon>0$，有

$$P(|X-E(X)|\geq\varepsilon)\leq\frac{D(X)}{\varepsilon^2},$$

即

$$P(|X-E(X)|<\varepsilon)\geq 1-\frac{D(X)}{\varepsilon^2}.$$

定理 1 (伯努利大数定律)**.** 在独立重复试验中，记事件 A 的概率为 p. 以 n_A 表示前 n 次试验中事件 A 发生的次数，则对任意正数 ε,

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1,$$

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\geq\varepsilon\right\}=0.$$

注记. 伯努利定理的实际意义：当重复试验次数充分大时，某事件发生的频率与该事件发生的概率有一定偏差的可能性很小.

证明. 因为 $n_A\sim b(n,p)$, 故

$$E(n_A)=np,\quad D(n_A)=np(1-p).$$

由数学期望和方差的性质, 有

$$E\left(\frac{n_A}{n}\right)=p, D\left(\frac{n_A}{n}\right)=\frac{1}{n^2}D(n_A)=\frac{p(1-p)}{n}.$$

任取 $\varepsilon>0$, 由切比雪夫不等式可得

$$P\left\{\left|\frac{n_A}{n}-p\right|\geqslant\varepsilon\right\}\leqslant\frac{1}{\varepsilon^2}D\left(\frac{n_A}{n}\right)=\frac{p(1-p)}{n\varepsilon^2}\to 0\quad(n\to\infty)$$

即

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1-\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n}-p\right|\geqslant\varepsilon\right\}=1.$$

若记 $X_i=\begin{cases}1, & \text{第}i\text{ 次试验中事件}A\text{ 出现}\\ 0, & \text{第}i\text{次试验中事件}A\text{ 不出现}\end{cases}$, 则

$$n_A=\sum_{i=1}^n X_i,\ \frac{n_A}{n}=\frac{1}{n}\sum_{i=1}^n X_i,\ p=\frac{1}{n}\sum_{i=1}^n P(A)=\frac{1}{n}\sum_{i=1}^n E(X_i).$$

定理 1 可写成

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{i=1}^n X_i-\frac{1}{n}\sum_{i=1}^n E(X_i)\right|<\varepsilon\right\}=1.$$

一般地, 若随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$ 的数学期望都存在, 且满足上式, 则称随机变量序列 $\{X_n\}$ 满足大数定律.

定理 2 (切比雪夫大数定律)**.** 设随机变量 $X_1,...,X_n,\cdots$ 相互独立，且有相同的期望 μ 和方差 σ^2，定义 Y_n 为前 n 个随机变量的算术平均，即

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,$$

则对任意正数 ε，有 $\lim\limits_{n\to\infty}P\left\{\left|\overline{X}-\mu\right|<\varepsilon\right\}=1.$

注记**.** 切比雪夫定理的实际意义：多次试验求平均值能够有效地控制误差.

证明. 证由于

$$E\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n}\sum_{i=1}^{n}E\left(X_k\right)=\frac{1}{n}n\mu=\mu$$

$$D\left(\frac{1}{n}\sum_{k=1}^{n}X_k\right)=\frac{1}{n^2}\sum_{k=1}^{n}D\left(X_k\right)=\frac{1}{n^2}n\sigma^2=\frac{\sigma^2}{n}$$

任取 $\varepsilon>0$, 由切比雪夫不等式可得

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}\geqslant 1-\frac{\sigma^2/n}{\varepsilon^2}$$

在上式中令 $n\to\infty$, 并注意概率不能大于 1 , 可得

$$\lim_{n\to x}P\left\{\left|\frac{1}{n}\sum_{k=1}^{\prime\prime}X_k-\mu\right|<\varepsilon\right\}=1.$$

定理 3 (辛钦大数定律)**.** 设随机变量 $X_1,...,X_n,\cdots$ 相互独立，且服从同一分布，具有数学期望 $E(X_k)=\mu$ $(k=1,2,\cdots)$, 则对任意正数 ε，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}=1.$$

注记**.** 辛钦大数定律不要求方差存在，但要求随机变量服从同一分布.

5.2 中心极限定理

中心极限定理的主要思想：如果

1. 一个随机现象由众多的随机因素所引起，
2. 且每一因素在总的变化里所起的作用不显著，

则描述这个随机现象的随机变量近似地服从正态分布.

这就是为什么实际中遇到的随机变量很多都服从正态分布的原因，也正因如此，正态分布在概率论和数理统计中占有极其重要的地位.

定理 **1** (林德伯格－莱维定理)**.** 若 $\{X_n\}_{n\geq 1}$ 独立且同分布，$E(X_i)=\mu$，$D(X_i)=\sigma^2>0$，令

$$Y_n=\frac{\sum_{i=1}^n X_i-E\left(\sum_{i=1}^n X_i\right)}{\sqrt{D\left(\sum_{i=1}^n X_i\right)}}=\frac{\sum\limits_{i=1}^n X_i-n\mu}{\sqrt{n}\sigma}$$

则其分布函数 $F_n(x)$ 收敛到 $\Phi(x)$，即对任何实数 x，

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

其中 $\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数，

定理 1 实际上说明了

$$\frac{\sum_{i=1}^n X_i-n\mu}{\sqrt{n}\sigma}\overset{\text{近似}}{\sim}N(0,1)$$

由此可见，当 n 充分大时

$$\sum_{i=1}^n X_i\overset{\text{近似}}{\sim}N\left(n\mu,n\sigma^2\right)$$

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

例 **1.** 已知产品的长度服从期望为 14、方差为 4 的分布. 求 100 件产品的平均长度超过 14.5 的概率.

例 **2.** 计算机进行加法计算时，把每个加数四舍五入变为整数来计算. 设所有的取整误差是相互独立的随机变量，并且都服从区间 $[-0.5,0.5)$ 上的均匀分布. 若独立进行了 300 次实数加法运算，求所有舍入误差的总和的绝对值小于 10 的概率.

练习 **1.** 一个螺丝钉的重量是一个随机变量，期望值是 10 克，标准差是 1 克，求一盒 (100 个) 同型号螺丝钉的重量超过 1020 克的概率.

设在某试验中事件 A 发生的概率为 p，将该试验独立地进行 n 次. 记 η_n 为 n 次试验中事件 A 发生的总次数，X_i 为第 i 次试验中事件 A 发生的次数，则

$$\eta_n \sim B(n,p),$$

$$X_i \sim B(1,p), i=1,2,\cdots,n,$$

且

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

易知：

$$E(\eta_n) = np,\ D(\eta_n) = np(1-p).$$

定理 **2** (棣莫弗一拉普拉斯定理)**.** 设随机变量 η_n 服从参数为 n,p 的二项分布，即 $\eta_n \sim B(n,p)$，则对任意 x 有

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

注记**.** 棣莫弗一拉普拉斯定理表明，当 n 充分大时，η_n 近似服从正态分布，即可以近似认为

$$\eta_n \sim N(np, np(1-p)).$$

或者

$$\frac{\eta_n - np}{\sqrt{np(1-p)}} \sim N(0,1).$$

由棣莫弗一拉普拉斯中心极限定理易知: 当 n 充分大时, 对任意 $a<b$, 有

$$\begin{aligned} P\{a \leqslant \eta_n \leqslant b\} &= P\left\{\frac{a-np}{\sqrt{np(1-p)}} \leqslant \frac{\eta_n - np}{\sqrt{np(1-p)}} \leqslant \frac{b-np}{\sqrt{np(1-p)}}\right\} \\ &\approx \Phi\left(\frac{b-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

例 **3.** 独立地掷 10 颗骰子, 求掷出的点数之和在 30 到 40 点之间的概率.

解. 以 X_i 表示第 i 颗骰子掷出的点数 $(i=1,2,\cdots,10)$, 则

$$P\{X_i=j\}=\frac{1}{6}, j=1,2,\cdots,6.$$

从而

$$E(X_i)=\mu=\frac{7}{2}, D(X_i)=\sigma^2=\frac{35}{12}$$

有中心极限定理可得

$$\begin{aligned}P\left\{30\leqslant\sum_{i=1}^{10}X_i\leqslant 40\right\}&=P\left\{\frac{30-10\times\frac{7}{2}}{\sqrt{10\times\frac{35}{12}}}\leqslant\frac{\sum_{i=1}^{10}X_i-10\times\frac{7}{2}}{\sqrt{10\times\frac{35}{12}}}\leqslant\frac{40-10\times\frac{7}{2}}{\sqrt{10\times\frac{35}{12}}}\right\}\\&\approx\Phi\left(\frac{40-35}{\sqrt{10\times\frac{35}{12}}}\right)-\Phi\left(\frac{30-35}{\sqrt{10\times\frac{35}{12}}}\right)\\&=2\Phi\left(\sqrt{\frac{6}{7}}\right)-1\approx 0.65.\end{aligned}$$

例 **4.** 在一家保险公司有一万人参加保险, 每年每人付 *12* 元保险费. 在一年内这些人死亡的概率都为 0.006, 死亡后家属可向保险公司领取 *1000* 元, 试求:

(1) 保险公司一年的利润不少于 *6* 万元的概率;

(2) 保险公司亏本的概率.

解. 设参加保险的一万人中一年内死亡的人数为 X, 则

$$X\sim b(10000,0.006).$$

由题设, 公司一年收人保险费 *12* 万元, 付给死者家属 $1000X$ 元, 于是, 公司一年的利润为

$$120000-1000X=1000(120-X).$$

由拉普拉斯中心极限定理开可得:

(1) 保险公司一年的利润不少于 6 万元的概率为

$$\begin{aligned}P\{1000(120-X)\geqslant 60000\}&=P\{0\leqslant X\leqslant 60\}\\&\approx\Phi\left(\frac{60-60}{7.72}\right)-\Phi\left(\frac{0-60}{7.72}\right)\\&=\Phi(0)-\Phi(-7.77)\approx 0.5-0=0.5.\end{aligned}$$

(2) 保险公司亏本的概率为

$$P\{1000(120-X)<0\}=P\{X>120\}=P\left\{\frac{X-60}{7.72}>\frac{120-60}{7.72}\right\}$$
$$\approx\frac{1}{\sqrt{2\pi}}\int_{7.77}^{\infty}\mathrm{e}^{-\frac{t^2}{2}}\,\mathrm{d}t$$
$$=1-\Phi(7.77)\approx 1-1=0.$$

例 **5.** 独立地测量一个物理量, 每次测量产生的误差都服从区间 $(-1,1)$ 上的均匀分布.

(1) 如果取 n 次测量的算术平均值作为测量结果, 求它与真值的差小于 ε 的概率;

(2) 计算 (1) 中当 $n=36,\varepsilon=\frac{1}{6}$ 时，概率的近似值;

(3) 取 $\varepsilon=\frac{1}{6}$, 要使上述概率不小于 $\alpha=0.95$, 应进行多少次测量?

解**.** 用 μ 表示所测量物理量的真值, X_i 表示第 i 次测量值, ε_i 表示第 i 次测量所产生的随机误差 $(i=1,2,\cdots,n)$, 于是 $X_i=\mu+\varepsilon_i$, 由题设 $\varepsilon_i\sim U(-1,1)$, 所以

$$E(\varepsilon_i)=0,\quad D(\varepsilon_i)=\frac{[1-(-1)]^2}{12}=\frac{1}{3},$$
$$E(X_i)=\mu,\quad D(X_i)=D(\mu+\varepsilon_i)=\frac{1}{3}\quad(i=1,2,\cdots,n).$$

又 $X_1,X_2,\cdots,X_n$ 独立同分布, 由中心极限定理, 当 n 充分大时, 随机变量

$$\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{\frac{n}{3}}}\sim N(0,1).$$

于是，(1) 中所求概率为

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=P\left\{\left|\sum_{i=1}^{n}X_i-n\mu\right|<n\varepsilon\right\}$$
$$=P\left\{\left|\frac{\sum_{i=1}^{n}X_i-n\mu}{\sqrt{\frac{n}{3}}}\right|<\varepsilon\sqrt{3n}\right\}$$
$$\approx 2\Phi(\sqrt{3n}\varepsilon)-1.$$

(2) 当 $n=36, \varepsilon=\frac{1}{6}$ 时, 所求概率为

$$
\begin{aligned}
P\left\{\left|\frac{1}{36}\sum_{i=1}^{36}X_i-\mu\right|<\frac{1}{6}\right\} &\approx 2\Phi\left(\frac{1}{6}\sqrt{3\times 36}\right)-1\\
&=2\Phi(\sqrt{3})-1\\
&\approx 2\Phi(1.73)-1\\
&=0.92.
\end{aligned}
$$

(3) 要求 n, 使得

$$
P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}\approx 2\Phi(\sqrt{3n}\varepsilon)-1\geqslant\alpha
$$

即 $\Phi(\sqrt{3n}\varepsilon)\geqslant\frac{1+\alpha}{2}$, 为此对给定的 α, 先查标准正态分布表求 λ, 使 $\Phi(\lambda)\geqslant\frac{1+\alpha}{2}$, 再令 $\varepsilon\sqrt{3n}\geqslant\lambda$, 由此确定出 n. 对于 $\alpha=0.95, \varepsilon=\frac{1}{6}$, 查表得 $\lambda=1.96$, 从而

$$
n\geqslant\frac{\lambda^2}{3\varepsilon^2}=\frac{(1.96)^2}{3\times\frac{1}{36}}\approx 46.
$$

练习 **2.** 某公司有 200 名员工参加一种资格证书考试. 按往年经验, 该考试的通过率为 0.8. 试计算这 200 名员工至少有 150 人考试通过的概率.

练习 **3.** 设电站供电网有 10000 盏电灯, 假设夜晚每一盏灯开灯概率都是 0.7, 而且各盏灯的开关时间彼此独立, 试估计夜晚同时开着的灯数在 6900 到 7100 之间的概率.

(1) 用切比雪夫不等式估计.

(2) 用中心极限定理估计.

练习 **4.** 扔硬币 100 个, 估计正面朝上的个数在 40 到 60 之间的概率.

(1) 用切比雪夫不等式估计.

(2) 用中心极限定理估计.