第二节　中心极限定理

一、列维—林德伯格中心极限定理

定理 1 设随机变量序列 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且

$EX_i = \mu$ ， $DX_i = \sigma^2 > 0$ ， $i = 1, 2, \cdots$ ． 令 $Y_n = \dfrac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma}$ ，

$n = 1, 2, \cdots$ ． Y_n 的分布函数记作 $F_{Y_n}(x)$，则有

$$\lim_{n\to\infty} F_{Y_n}(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt，\quad x \in (-\infty, +\infty) .$$

注 1： 由定理 1 表明，当 n 充分大时，,

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

从而有

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

即

$$F_{Y_n}(x) \approx \Phi(x)$$

注 2： 特别地，当 $n \geq 30$ 时，其误差可以忽略不计.

注 3： 由于定理 1 中的 X_i 的概率分布可以是任意的，因此，$\sum_{i=1}^{n} X_i$ 的概率分布难于精确求得．但只要 n 充分大，则有 $\sum_{i=1}^{n} X_i$ 近似服从正态分布，因而突出了正态分布在概率统计中的重要地位．

在实际问题中，有些随机变量是诸多独立同分布，且影响甚微的小因素叠加而成的，因此这些随机变量可近似刻画成服从正态分布的随机变量，这就是中心极限定理的客观背景．

例 1 设随机变量 $X_1, X_2, \cdots, X_{48}$ 相互独立，且

$$X_i \sim U(0,2), i=1,2,\cdots,48.$$

记 $X=\sum_{i=1}^{48} X_i$，则利用中心极限定理计算 $P\{X \le 50\}$.

解 由于 $EX_i=1, DX_i=\frac{1}{3}, \ i=1,2,\cdots,48$，所以 $EX=48,$ $DX=48\times\frac{1}{3}=16$．由中心极限定理知

$$X=\sum_{i=1}^{48} X_i \overset{\text{近似}}{\sim} N(48,16),$$

故 $P\{X \le 50\}=P\{\frac{X-48}{4} \le \frac{1}{2}\}=\Phi(\frac{1}{2})=0.6915$.

例 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，才能保障不超载的概率大于 0.977．

解 设每辆车可以装 n 箱．记 X_i 为第 i 箱的重量（单位：千克），$i=1,2,\cdots,n$，由题意知 $X_1,X_2,\cdots,X_n$ 为独立同分布的随机变量，并且 $EX_i=50, DX_i=25$， $i=1,2,\cdots,n$．

（续解）根据列维–林德伯格中心极限定理，知总重量

$$T_n=\sum_{i=1}^{n}X_i \overset{\text{近似}}{\sim} N(50n,25n).$$

由题意知，不超载的概率

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

$$=\Phi\left(\frac{1000-10n}{\sqrt{n}}\right)>0.977=\Phi(2).$$

由此可见，$\dfrac{1000-10n}{\sqrt{n}}>2$，从而$n<98.0199$，即最多可以装98箱.

二、棣莫弗—拉普拉斯中心极限定理

定理 2 设随机变量 $X_n \sim B(n,p)$， $n=1,2,\cdots$， 则

$$\lim_{n\to\infty} P\left\{ \frac{X_n - np}{\sqrt{np(1-p)}} \le x \right\} = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt ， \quad x \in R .$$

注 4： 定理 2 的应用：若 $X \sim B(n,p)$，则当 n 充分大时，有

$$X \overset{\text{近似}}{\sim} N\left(EX, DX\right) = N(np, np(1-p)) .$$ （主要结论）

同样，当 $n \ge 30$ 时，其误差可以忽略不计.

定理 2 称为棣莫弗—拉普拉斯中心极限定理，也称为二项分布以正态分布为极限分布的中心极限定理.

例 3 设有 200 台独立作业的同类型设备，每台设备出现故障的概率均为 $\frac{1}{200}$，又设每台设备的故障可由一名维修人员处理．问至少配备多少名维修人员，才能使出现故障而不能即时维修的概率不大于 0.03？

解 设 X 表示故障设备台数，则 $X \sim B(200, \frac{1}{200})$， 由定理 2，

$$X \overset{\text{近似}}{\sim} N(1, \frac{199}{200}).$$

设 n 为需配备的维修人员的人数，由 $P\{X > n\} \le 0.03$，

得 $P\{X \le n\} \ge 0.97$，计算得 $\Phi(\frac{n-1}{\sqrt{199/200}}) \ge 0.97 = \Phi(1.88)$，

从而有 $\frac{n-1}{\sqrt{199/200}} \ge 1.88$，解得 $n \ge 2.88$，所以 n 至少取 3，即至少配备 3 名维修人员.

练习：

1. 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且均服从参数为 $\lambda(\lambda>0)$ 的指数分布，则（ ）.

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

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

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

答案：(D).