第五章　极限定理

极限定理是概率论的基本理论之一，在概率论和数理统计的理论研究和实际应用中都具有重要的意义．在这一章，我们将介绍有关随机变量序列的最基本的两类极限定理，即大数定律和中心极限定理．

注意到，随机现象的统计规律性是在相同条件下进行大量重复试验时呈现出来的．例如，在概率的统计定义中，谈到一个事件发生的频率具有稳定性，即频率趋于事件的概率．这里是指试验的次数无限增大时，在某种收敛意义下逼近某一定数．这就是最早的一个大数定律．一般的大数定律讨论 n 个随机变量的平均值的稳定性．大数定律对上述情况从理论的高度给予概括和论证．

另一类基本的极限定理是中心极限定理．这些定理证明了，在很一般的条件下，n 个随机变量的和当 $n\to\infty$ 时的极限分布是正态分布．利用这些结论，在数理统计中许多复杂随机变量的分布可以用正态分布近似，而正态分布有许多完美的理论，从而可以获得既实用又简单的统计分析．

本章将介绍大数定律和中心极限定理中的最简单也是最重要的结论．

§5.1　大数定律

一、切比雪夫(Chebyshev)不等式

首先介绍一个重要的不等式——切比雪夫不等式．

定理 5.1.1　设随机变量 X 具有期望 $E(X)=\mu$，方差 $\mathrm{Var}(X)=\sigma^2$，则对于任意正数 ε，有

$$P\{|X-\mu|\geqslant\varepsilon\}\leqslant\frac{\sigma^2}{\varepsilon^2}. \tag{5.1.1}$$

证　只对 X 是连续型随机变量的情况加以证明．设 X 的概率密度函数为 $f(x)$，则有(参见图 5.1.1)

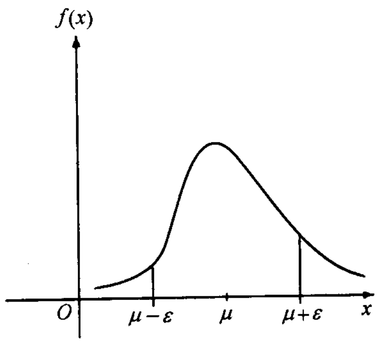

图 5.1.1　(5.1.1)式的证明

$$P\{|X-\mu|\geqslant\varepsilon\}$$

$$=\int_{|x-\mu|\geqslant\varepsilon}f(x)\mathrm{d}x$$

$$\leqslant \int_{|x-\mu|\geqslant\varepsilon} \frac{|x-\mu|^2}{\varepsilon^2} f(x)\mathrm{d}x$$

$$\leqslant \frac{1}{\varepsilon^2}\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\mathrm{d}x$$

$$= \frac{\sigma^2}{\varepsilon^2}.$$

定理证毕.

切比雪夫不等式也可以写为

$$P\{|X-\mu|<\varepsilon\}\geqslant 1-\frac{\sigma^2}{\varepsilon^2}. \tag{5.1.2}$$

切比雪夫不等式说明,若 X 的方差小,则事件$\{|X-\mu|<\varepsilon\}$发生的概率就大,即 X 取的值基本上集中于它的期望 μ 附近.这进一步说明了方差的意义.

用切比雪夫不等式可以在 X 的分布未知的情况下,估计概率值 $P\{|X-\mu|<\varepsilon\}$或 $P\{|X-\mu|\geqslant\varepsilon\}$,例如,

$$P\{|X-\mu|<3\sigma\}\geqslant 1-\frac{\sigma^2}{9\sigma^2}=0.8889.$$

此外,切比雪夫不等式作为一个理论工具,它的应用是普遍的,在大数定律的证明中将要用到它.

二、大数定律

首先引入随机变量序列 $X_1,X_2,\cdots,X_n,\cdots$相互独立的概念.如果对于任意 $n>1$, $X_1,X_2,\cdots,X_n$ 相互独立,则称 $X_1,X_2,\cdots,X_n,\cdots$相互独立.

定理 5.1.2　设随机变量 $X_1,X_2,\cdots,X_n\cdots$相互独立,并且具有相同的期望和方差:$E(X_i)=\mu,\mathrm{Var}(X_i)=\sigma^2,i=1,2,\cdots$.作前 n 个随机变量的平均 $Y_n=\frac{1}{n}\sum_{i=1}^{n}X_i$,则对于任意正数 ε,有

$$\lim_{n\to\infty}P\{|Y_n-\mu|<\varepsilon\}=1. \tag{5.1.3}$$

证

$$E(Y_n)=\frac{1}{n}\sum_{i=1}^{n}E(X_i)=\mu,$$

$$\mathrm{Var}(Y_n)=\frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(X_i)=\frac{\sigma^2}{n},$$

使用切比雪夫不等式,得到

$$P\{|Y_n-\mu|<\varepsilon\}\geqslant 1-\frac{\sigma^2}{n\varepsilon^2},$$

令 $n\to\infty$,注意到概率不可能大于 1,得到(5.1.3)式. 定理证毕.

(5.1.3)式表明,对于任意正数 ε,当 n 很大时,事件$\{|Y_n-\mu|<\varepsilon\}$发生的概率很大,从概率意义上指出了当 n 很大时,Y_n 接近 μ 的确切含义. 在概率论中,把(5.1.3)式表示的收敛性称为随机变量序列 $Y_1,Y_2,\cdots,Y_n,\cdots$依概率收敛于 μ,记为 $Y_n\xrightarrow{P}\mu$.

定理 5.1.2 表明,当 n 很大时,随机变量 $X_1,X_2,\cdots,X_n$ 的平均值 Y_n 接近于期望 μ,这种接近是在概率意义下的接近. 也就是说,不论给定怎样小的 $\varepsilon>0$,Y_n 与 μ 的偏离大于等于 ε 是可能的,但是当 n 很大时,出现这种偏离的可能性很小. 因此,当 n 很大时,我们有很大的把握保证 Y_n 很接近 μ.

推论 设 n 次独立重复试验中事件 A 发生的次数为 n_A,在每次试验中事件 A 发生的概率为 p,则对于任意正数 ε,有

$$\lim_{n\to\infty}P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1. \tag{5.1.4}$$

证 令

$$X_i=\begin{cases}1, & \text{若在第 } i \text{ 次试验中事件} A \text{ 发生},\\ 0, & \text{若在第 } i \text{ 次试验中事件} A \text{ 不发生},\end{cases}$$

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

则 $X_1,X_2,\cdots$是相互独立的随机变量序列. 显然 X_i 服从两点分布,从而

$$E(X_i)=p,\ \mathrm{Var}(X_i)=p(1-p),\ i=1,2,\cdots.$$

注意到

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

由定理 5.1.2 得到(5.1.4)式.

这个推论就是最早的一个大数定律,称为伯努利定理. 该定理表明事件 A 发生的频率$\frac{n_A}{n}$依概率收敛于事件 A 的概率 p,以严格的数学形式表达了频率的稳定性,即随着试验次数的增加,事件发生的频率逐渐稳定于事件的概率. 这个事实为我们在实际应用中用频率去估计概率提供了一个理论依据.

§5.2 中心极限定理

中心极限定理是棣莫佛(De Moivre)在18世纪首先提出的,到现在内容已十分丰富.在这一节里,我们只介绍其中两个最基本的结论.

定理 5.2.1 (独立同分布的中心极限定理) 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立,服从同一分布,并且具有期望和方差:$E(X_i)=\mu$, $\mathrm{Var}(X_i)=\sigma^2>0, i=1,2,\cdots$,则随机变量

$$Y_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\sigma} \tag{5.2.1}$$

的分布函数 $F_n(x)$ 收敛到标准正态分布函数,即对于任意实数 x 满足

$$\lim_{n\to\infty} F_n(x) = \lim_{n\to\infty} P\{Y_n \leqslant x\} = \Phi(x), \tag{5.2.2}$$

其中 $\Phi(x)$ 是标准正态分布 $N(0,1)$ 的分布函数,即

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

定理 5.2.1 的证明从略.这个定理的证明是20世纪20年代由林德伯格(Lindeberg)和莱维(Levy)给出的.

(5.2.2)式表明,随机变量序列 $Y_1, Y_2, \cdots, Y_n, \cdots$ 的分布函数序列 $\{F_n(x)\}$ 的极限是 $\Phi(x)$.

注意到 $E(\sum_{i=1}^{n} X_i) = n\mu$, $\mathrm{Var}(\sum_{i=1}^{n} X_i) = n\sigma^2$,(5.2.1)式可写为

$$Y_n = \frac{\sum_{i=1}^{n} X_i - E(\sum_{i=1}^{n} X_i)}{\sqrt{\mathrm{Var}(\sum_{i=1}^{n} X_i)}}, \tag{5.2.3}$$

从而 Y_n 的期望是 0,方差是 1, Y_n 是 $\sum_{i=1}^{n} X_i$ 的标准化的随机变量.由定理 5.2.1,当 n 很大时, Y_n 近似服从标准正态分布 $N(0,1)$,从而当 n 很大时, $\sum_{i=1}^{n} X_i$ 近似服从正态分布 $N(n\mu, n\sigma^2)$.由于 X_i 的分布在一定程度上可以是任意的,一般说来, $\sum_{i=1}^{n} X_i$ 的分布难于确切求得,这时,只要 n 很大,就能通过 $\Phi(x)$ 给出 $\sum_{i=1}^{n} X_i$ 的分布函数的近似值.这是正态分布在概率统计中占有重要

地位的一个基本原因.

在实际问题中,有很多随机现象可以看作是许多因素的独立影响的综合结果,而每一因素对该现象的影响都很微小,那么,描述这种随机现象的随机变量可以看成许多相互独立的起微小作用的因素的总和,它往往近似服从正态分布.这就是中心极限定理的客观背景.例如,测量误差是由许多观察不到的、可加的微小误差所合成的;在任一指定时刻,一个城市的耗电量是大量用户耗电量的总和等,它们都可以用服从正态分布的随机变量近似地描述.

在数理统计中,中心极限定理是大样本统计推断的理论基础.

例 5.2.1 设一批产品的强度服从期望为 14,方差为 4 的分布.每箱中装有这种产品 100 件,问:

(1) 每箱产品的平均强度超过 14.5 的概率是多少?

(2) 每箱产品的平均强度超过期望 14 的概率是多少?

解 $n=100$.设 X_i 是第 i 件产品的强度,$E(X_i)=14,\mathrm{Var}(X_i)=4,i=1,2,\cdots,100$.记 $\overline{X}=\frac{1}{100}\sum\limits_{i=1}^{100}X_i$,依定理 5.2.1,近似地有

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\frac{\overline{X}-14}{2/\sqrt{100}}=\frac{\overline{X}-14}{0.2}\sim N(0,1),$$

于是

$$\begin{aligned}(1)\ P\{\overline{X}>14.5\}&=P\left\{\frac{\overline{X}-14}{0.2}>\frac{14.5-14}{0.2}\right\}\\&=P\left\{\frac{\overline{X}-14}{0.2}>2.5\right\}\\&\approx1-\Phi(2.5)\\&=1-0.9938\\&=0.0062,\end{aligned}$$

可见,100 件产品的平均强度超过 14.5 的概率非常之小.

$$\begin{aligned}(2)\ P\{\overline{X}>14\}&=P\left\{\frac{\overline{X}-14}{0.2}>0\right\}\\&\approx\Phi(0)\\&=0.5,\end{aligned}$$

于是,我们可以说,每箱产品的平均强度超过 14 的概率约为 50%.

例 5.2.2 计算机在进行数字计算时,遵从四舍五入原则.为简单计,现在对小数点后面第一位进行舍入运算,则误差 X 可以认为服从$[-0.5,0.5]$

上的均匀分布.若在一项计算中进行了 100 次数字计算,求平均误差落在区间 $\left[-\frac{\sqrt{3}}{20}, \frac{\sqrt{3}}{20}\right]$上的概率.

解　$n=100, X_1, X_2, \cdots, X_{100}$ 相互独立,都服从$[-0.5, 0.5]$上的均匀分布.这时 $E(X_i)=0, \operatorname{Var}(X_i)=\frac{1}{12}, i=1,2,\cdots,100$,从而,近似地有

$$Y_{100}=\frac{\sum_{i=1}^{100} X_i-100 \times 0}{\sqrt{100/12}}=\frac{\sqrt{3}}{5} \sum_{i=1}^{100} X_i \sim N(0,1).$$

于是,平均误差 $\overline{X}=\frac{1}{100} \sum_{i=1}^{100} X_i$ 落在区间$\left[-\frac{\sqrt{3}}{20}, \frac{\sqrt{3}}{20}\right]$上的概率为

$$\begin{aligned} P\left\{-\frac{\sqrt{3}}{20} \leqslant \overline{X} \leqslant \frac{\sqrt{3}}{20}\right\} &= P\left\{-\frac{\sqrt{3}}{20} \leqslant \frac{1}{100} \sum_{i=1}^{100} X_i \leqslant \frac{\sqrt{3}}{20}\right\} \\ &= P\left\{-3 \leqslant \frac{\sqrt{3}}{5} \sum_{i=1}^{100} X_i \leqslant 3\right\} \\ &\approx \Phi(3)-\Phi(-3) \\ &= 0.9973. \end{aligned}$$

注意到 $\frac{\sqrt{3}}{20}=0.0866$,于是,平均误差 $\overline{X}$ 几乎取值于区间$[-0.0866, 0.0866]$之内.

下面介绍定理 5.2.1 的一个重要特例,是历史上最早的中心极限定理,由棣莫佛提出,拉普拉斯(Laplace)推广的,称为棣莫佛－拉普拉斯定理.

定理 5.2.2(棣莫佛－拉普拉斯定理)　设随机变量 $X_1, X_2, \cdots, X_n, \cdots$,相互独立,并且都服从参数为 p 的两点分布,则对于任意实数 x,有

$$\lim_{n \to \infty} P\left\{\frac{\sum_{i=1}^{n} X_i-n p}{\sqrt{n p(1-p)}} \leqslant x\right\}=\Phi(x). \tag{5.2.4}$$

证　$E(X_i)=p, \operatorname{Var}(X_i)=p(1-p), i=1,2,\cdots$.由(5.2.2)式即可得定理之结论.

由于 $\sum_{i=1}^{n} X_i \sim B(n, p)$,从而,定理 5.2.2 表明,有关二项分布的概率计算,当 n 很大时,可以通过关于 $\Phi(x)$的计算解决.另外,当 n 很大时,二项分布可以用正态分布来近似.

例 5.2.3 某单位有 200 部电话,每部电话约有 5% 的时间要使用外线通话.设每部电话是否使用外线通话是相互独立的,问该单位总机至少需要安装多少条外线,才能以 90% 以上的概率保证每部电话需要使用外线通话时可以打通?

解 令

$$X_i=\begin{cases}1, & \text{若第 } i \text{ 部电话使用外线通话},\\ 0, & \text{若第 } i \text{ 部电话不使用外线通话},\end{cases}$$

$$i=1,2,\cdots,200,$$

于是 $p=P\{X_i=1\}=0.05, np=10, np(1-p)=9.5$. $\sum_{i=1}^{200}X_i$ 是同时使用外线通话的电话总数,则问题为:求最小的 k 值,使得

$$P\{0\leqslant\sum_{i=1}^{200}X_i\leqslant k\}\geqslant 0.90.$$

使用定理 5.2.2 的结论,有

$$\begin{aligned}P\{0\leqslant\sum_{i=1}^{200}X_i\leqslant k\}&=P\{\frac{0-10}{\sqrt{9.5}}\leqslant\frac{\sum_{i=1}^{200}X_i-10}{\sqrt{9.5}}\leqslant\frac{k-10}{\sqrt{9.5}}\}\\&\approx\Phi(\frac{k-10}{\sqrt{9.5}})-\Phi(\frac{-10}{\sqrt{9.5}})\\&\geqslant 0.90.\end{aligned}$$

由于 $\Phi(\frac{-10}{\sqrt{9.5}})\approx 0$,于是只要 $\Phi(\frac{k-10}{\sqrt{9.5}})\geqslant 0.90$,即 $\frac{k-10}{\sqrt{9.5}}\geqslant 1.282$, $k\geqslant 13.95$,即至少安装 14 条外线.

例 5.2.4 某市保险公司开办一年人身保险业务,被保险人每年需交付保险费 160 元,若一年内发生重大人身事故,其本人或家属可获 2 万元赔金.已知该市人员一年内发生重大人身事故的概率为 0.005,现有 5000 人参加此项保险,问保险公司一年内从此项业务所得到的总收益在 20 万到 40 万元之间的概率是多少?

解 记

$$X_i=\begin{cases}1, & \text{若第 } i \text{ 个被保险人发生重大事故},\\ 0, & \text{若第 } i \text{ 个被保险人未发生重大事故},\end{cases}$$

$$i=1,2,\cdots,5000,$$

于是 X_i 均服从参数为 $p=0.005$ 的两点分布，$P(X_i=1)=0.005$，$np=25$. $\sum_{i=1}^{5000}X_i$ 是 5000 个被保险人中一年内发生重大人身事故的人数，保险公司一年内从此项业务所得到的总收益为 $0.016\times5000-2\times\sum_{i=1}^{5000}X_i$ 万元. 于是

$$
\begin{aligned}
&P\{20\leqslant 0.016\times5000-2\sum_{i=1}^{5000}X_i\leqslant 40\}\\
&=P\{20\leqslant\sum_{i=1}^{5000}X_i\leqslant 30\}\\
&=P\left\{\frac{20-25}{\sqrt{25\times0.995}}\leqslant\frac{\sum_{i=1}^{5000}X_i-25}{\sqrt{25\times0.995}}\leqslant\frac{30-25}{\sqrt{25\times0.995}}\right\}\\
&\approx\Phi(1)-\Phi(-1)\\
&=0.6826.
\end{aligned}
$$

习　题　五

5.1　已知正常男性成人每毫升的血液中，含白细胞平均数是 7300，均方差是 700. 使用切比雪夫不等式估计每毫升血液中含白细胞数在 5200 到 9400 之间的概率.

5.2　设随机变量 X 服从参数为 λ 的泊松分布，使用切比雪夫不等式证明

$$P\{0<X<2\lambda\}\geqslant\frac{\lambda-1}{\lambda}.$$

5.3　设由机器包装的每包大米的重量是一个随机变量，期望是 10 公斤，方差是 0.1 公斤^2. 求 100 袋这种大米的总重量在 990 至 1010 公斤之间的概率.

5.4　一加法器同时收到 20 个噪声电压 $V_i, i=1,2,\cdots,20$，设它们是相互独立的随机变量，并且都服从区间$[0,10]$上的均匀分布. 记 $V=\sum_{i=1}^{20}V_i$，计算 $P\{V>105\}$ 的近似值.

5.5　一复杂的系统由 100 个相互独立起作用的部件所组成，在整个运行期间每个部件损坏的概率为 0.1，为了使整个系统起作用，至少必须有 85 个部件正常工作，求整个系统起作用的概率.

5.6　银行为支付某日即将到期的债券需准备一笔现金. 设这批债券共发放了 500 张，每张债券到期之日需付本息 1000 元. 若持券人(一人一券)于债券到期之日到银行领取本息的概率为 0.4，问银行于该日应至少准备多少现金才能以 99.9% 的把握满足持券人的兑换?