第五章　大数定律及中心极限定理

北京化工大学数学系

苏贵福

极限定理是概率论的基本理论, 在理论研究和应用中起着重要的作用, 其中最重要的 是称为"大数定律"与"中心极限定理"的一些定理.

- 大数定律是叙述随机变量序列的前n项 的算术平均值在某种条件下收敛到这n项的均值的算术平均值.
- 中心极限定理则是确定在什么条件下, 大量随机变量之和的分布逼近于正态分布.

5.1大数定律

在第一章我们了解到: 当随机试验的次数n充分大时, 随机事件A发生的概率总是在一个常数p $(0 \le p \le 1)$附近波动.

在实际测量时, 为了提高测量精度, 往往进行多次测量, 然后用测量的实测值的平均值近似代替真值. 这正是大数定律的实际背景.

经验表明, 只要测量的次数足够多, 总可以达到事先要求的精度.

一. 切比雪夫不等式的意义

① 切比雪夫不等式

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

给出了随机变量X的取值落在以其均值$E(X)$为中心，以ϵ为半径的区域内的概率的一个上界估计，通常将此估计称为双侧尾概率估计.

② 切比雪夫不等式的优势在于它并不依赖于随机变量X的具体概率分布，因而有比较宽泛的使用范围. 但这一估计的精度不高.

例1 设$E(X)=\mu$, $D(X)=\sigma^2$. 由切比雪夫不等式可得

$$P\{|X-\mu|<2\sigma\} \geq 1-\frac{\sigma^2}{4\sigma^2}=0.75$$
$$P\{|X-\mu|<3\sigma\} \geq 1-\frac{\sigma^2}{9\sigma^2}=0.89.$$

如果设$X\sim N(\mu,\sigma^2)$, 则有

$$P\{|X-\mu|<2\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<2\right\}=\Phi(2)-\Phi(-2)=0.955$$
$$P\{|X-\mu|<3\sigma\}=P\left\{\left|\frac{X-\mu}{\sigma}\right|<3\right\}=\Phi(3)-\Phi(-3)=0.997.$$

♠ 由此说明当随机变量X的概率分布已知时, 双侧尾概率估计的精度大大提高.

二. 经典大数定律

定义1 设$\{X_n\}$是独立同分布的随机变量序列, 如果对任意的$\epsilon > 0$, 恒有

$$\lim_{n\to\infty} P\{|\overline{X} - \mu| \geq \epsilon\} = 0.$$

其中$\overline{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$, $\mu = E(X_n)$, 则称随机变量序列 $\{X_n\}$ 服从大数定律.

定理1 设随机变量$X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望和方差，$E(X_k) = \mu$，$D(X_k) = \sigma^2$，$(k = 1, 2, \cdots)$. 则对于任意给定的正数ϵ，都有

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

证明 令$Y_n = \frac{1}{n}\sum_{k=1}^{n} X_k$，则

$$\begin{aligned} E(Y_n) &= \frac{1}{n}[E(X_1) + E(X_2) + \cdots + E(X_n)] \\ &= \frac{1}{n} \cdot n\mu = \mu. \end{aligned}$$

又因为$X_1, X_2, \cdots, X_n, \cdots$独立，且$D(X_k) = \sigma^2$. 故

$$D(Y_n) = \frac{1}{n^2}[D(X_1) + D(X_2) + \cdots + D(X_n)] = \frac{\sigma^2}{n}.$$

由切比雪夫不等式得

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

也就是说

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \epsilon\right\} \geq 1 - \frac{\sigma^2}{n\epsilon^2}.$$

根据概率的性质有

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \epsilon\right\} \leq 1.$$

由两边夹定理有

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

定理2 (切比雪夫大数定律) 设随机变量$X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 每个变量分别存在数学期望和方差分别为

$$E(X_1), E(X_2), \cdots, E(X_n), \cdots$$

$$D(X_1), D(X_2), \cdots .D(X_n), \cdots$$

并且这些方差都是有界的. 则对于任意给定的正数ϵ, 都有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n}\sum_{k=1}^{n} E(X_k)\right| < \epsilon\right\} = 1.$$

① 切比雪夫大数定律是对定理1的一个推广.

② 切比雪夫大数定律表明: 当n充分大时, n个随机变量 的算术平均值$\frac{1}{n}\sum_{k=1}^{n} X_k$偏离其数学期望$\mu = \frac{1}{n}\sum_{k=1}^{n} E(X_k)$ 可能性很小.

③ 实际应用中往往用某物体的某个指标的一系列实测值的算术平均值近似代替其真值.

定理3 (贝努利大数定律) 设在n次重复独立试验 中事件A发生Y_n次, 每次试验A发生的概率为p, 则对任意的正数ϵ, 都有

$$\lim_{n\to\infty} P\left\{\left|\frac{Y_n}{n}-p\right|<\epsilon\right\}=1.$$

证明 令 $X_k=\begin{cases}0, & 第k次试验A不发生\\ 1, & 第k次试验A发生\end{cases}, k=1,2,\cdots.$

显然$Y_n=X_1+X_2+\cdots+X_n$. 因为X_k只依赖于第k次试验, 而各次试验是独立的, 所以$X_1,X_2,\cdots,X_n$相互独立. 又因为X_k服从参数为p的$(0-1)$分布, 故$E(X_k)=p$, $D(X_k)=p(1-p)$, $k=1,2,\cdots$. 于是由定理1得

$$\lim_{n\to\infty} P\left\{\left|\frac{Y_n}{n}-p\right|<\epsilon\right\}=1.$$

贝努利大数定理以严格的数学形式诠释了频率的稳定性. 即当n很大时, 事件发生的频率与概率有较大偏差的可能性很小.

因此在实际应用中, 当试验的次数很大时, 便可以用事件发生的频率近似代替事件发生的概率.

定理4 (辛钦大数定律) 设$X_1, X_2, \cdots, X_n, \cdots$是相互独立, 服从同一分布的随机变量序列, 且具有数学期望$E(X_k) = \mu$, $k = 1, 2, \cdots$. 则对任意的正数ϵ, 都有

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