Chap5.大数定律与中心极限定理

44.大数定律

随机变量序列的极限（依概率收敛）

设 $\{X_n\}, n=1,2,3,...$ 是一随机变量序列，a 是一个常数。若对于任意正数 $\epsilon>0$，有

$$\lim_{n\to\infty} P\left(|X_n-a|<\epsilon\right)=1, \tag{1}$$

则称随机变量序列 $\{X_n\}$ 依概率收敛于 a，记为 $X_n \xrightarrow{P} a$。

性质：若 $X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$，二元连续函数 $g(x,y)$，则 $g(X_n,Y_n)\to g(a,b)$。

[注] 在讨论未知参数估计量是否具有一致性（相合性）时，常常用到依概率收敛的这一性质和大数定律。

> **依概率收敛**
>
> 通俗来说，依概率收敛就是指在一系列随机试验中，随着试验次数的增加，某个变量的值会越来越接近与一个固定的值。
>
> 例如：抛硬币，虽然过程是随机的，但是随着次数的增加，结果会越来越接近于这个期望值。这个过程就可以叫做依概率收敛。

虽然每次实验的具体结果不一样，但随着实验次数的增加，随机变量的结果会越来越多地集中在 a 附近。

弱大数定理（辛钦大数定理）

设随机变量 $X_1,X_2,\ldots,X_n,\ldots$ 相互独立，服从一分布，且具有数学期望 $E(X_k)=\mu\ (k=1,2,\ldots)$，作前 n 个随机变量的算术平均 $\frac{1}{n}\sum_{k=1}^{n}X_k$，则对任意 $\epsilon>0$，有：

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\epsilon\right)=1, \tag{2}$$

或者

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|\geq\epsilon\right)=0. \tag{3}$$

回忆：切比雪夫不等式，通过切比雪夫不等式证明

$$P\left(|X-E(X)|\leq\epsilon\right)\geq 1-\frac{\sigma^2}{\epsilon^2}. \tag{4}$$

> **证明**：假设 $X_1, X_2, ...$ 相互独立，且 $E(X_k) = \mu$。
>
> - $E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}(E(X_1)+E(X_2)+\cdots+E(X_n)) = \frac{n\mu}{n} = \mu$。
> - $D\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}D(X_1+X_2+\cdots+X_n) = \frac{1}{n^2}(D(X_1)+D(X_2)+\cdots+D(X_n))$。
>
> 由切比雪夫不等式：
>
> $$P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \epsilon\right) \leq \frac{\sigma^2}{n\epsilon^2}. \tag{5}$$
>
> 当 $n \to \infty$ 时，$\frac{\sigma^2}{n\epsilon^2} \to 0$，因此：
>
> $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \geq \epsilon\right) = 0. \tag{6}$$
>
> 由此得出弱大数定理，$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$。
>
> **注**：
>
> $$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| \leq \epsilon\right) = 1 \tag{7}$$
>
> 或者
>
> $$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu = E(X_k) \tag{8}$$

例如，设 $X_1, X_2, \ldots, X_n, \ldots$ 为相互独立且服从从参数 2 指定的分布，则当 n 趋大时，$\frac{1}{n}\sum_{k=1}^{n} X_k$ 后概率收敛于 $\mu = 2$。

辛钦大数定理的意思是：

如果我们进行大量独立同分布的随机实验，那么这些实验结果的平均值会越来越接近数学期望。

例如，假设抛骰子，做大量独立重复的试验，那么这个游戏的结果会逐渐接近与骰子的期望值 3.5

伯努利大数定理

设 f_A 是 n 次独立重复试验中事件 A 发生的次数，p 是事件 A 在每次试验中发生的概率，则对于任意正数 $\epsilon > 0$，有：

$$\lim_{n\to\infty} P\left(\left|\frac{f_A}{n} - p\right| < \epsilon\right) = 1, \tag{9}$$

或

$$\lim_{n\to\infty} P\left(\left|\frac{f_A}{n} - p\right| \geq \epsilon\right) = 0. \tag{10}$$

> **证明**：
> f_A 是 n 次独立试验中事件 A 发生的次数，$p = P(A)$。
>
> 令 X_k 为：
>
> $$X_k = \begin{cases} 1, & \text{第 k 次试验中事件 A 发生} \\ 0, & \text{第 k 次试验中事件 A 不发生} \end{cases} \tag{11}$$
>
> $E(X_k) = P(A)$。
>
> 假设 $X_1, X_2, \ldots, X_n$ 相互独立且同分布，服从 $0-1$ 分布，且
>
> $$f_A = X_1 + X_2 + \cdots + X_n. \tag{12}$$
>
> 除以 n 得：
>
> $$\frac{f_A}{n} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n), \tag{13}$$
>
> 由大数定理，得：
>
> $$\frac{f_A}{n} = \frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} P(A). \tag{14}$$
>
> 因此，
>
> $$\lim_{n\to\infty} P\left(\left|\frac{f_A}{n} - P(A)\right| < \epsilon\right) = 1. \tag{15}$$
>
> **注**：
>
> - 频率：$\frac{f_A}{n} \to P(A)$ 代表概率。
> - 当 n 趋大时，事件 $|\frac{f_A}{n} - P(A)| < \epsilon$ 发生的概率接近 1。

即：当试验次数趋于无穷时，事件的频率与概率 $P(A)$ 偏差趋于 0，发生频率趋近于理论概率。

应用：在实际应用中，当试验次数非常大时，可以用事件的发生频率来近似概率。

45.中心极限定理

独立同分布的中心极限定理

设随机变量 $X_1, X_2, \ldots, X_n$ 满足以下条件：

1. 相互独立，
2. 服从同一分布，

3. 具有数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2\ (k=1,2,\ldots)$，

则随机变量之和 $\sum_{i=1}^{n} X_i$ 的标准化变量：

$$Y_n=\frac{\sum_{k=1}^{n} X_k-E\left(\sum_{k=1}^{n} X_k\right)}{\sqrt{D\left(\sum_{k=1}^{n} X_k\right)}}=\frac{\sum_{k=1}^{n} X_k-n\mu}{\sqrt{n\sigma^2}} \tag{16}$$

的分布函数 $F_n(x)$ 对于任意 $\epsilon>0$ 满足：

$$\lim_{n\to\infty} F_n(x)=\lim_{n\to\infty} P\left(\frac{\sum_{k=1}^{n} X_k-n\mu}{\sqrt{n\sigma^2}}\leq x\right) \tag{17}$$

即：

$$=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt=\Phi(x) \tag{18}$$

其中，$\Phi(x)$ 为标准正态分布的分布函数。

棣莫弗-拉普拉斯定理

设随机变量 $\eta_n\ (n=1,2,\ldots)$ 服从参数为 $n,p\ (0<p<1)$ 的二项分布，则对于任意 x，有：

$$\lim_{n\to\infty} P\left(\frac{\eta_n-np}{\sqrt{np(1-p)}}\leq x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}dt=\Phi(x) \tag{19}$$

其中，$\Phi(x)$ 为标准正态分布的分布函数。

棣莫弗-拉普拉斯定理表明，在进行大量的独立二项试验时，二项分布会近似于一个正态分布。

这意味着，即使试验的结果是离散的（二项分布是离散型分布），当样本量 n 足够大时，也可以使用正态分布的连续性来简化计算和推断。

设随机变量 X 服从二项分布 $X\sim B(n,p)$，即 X 表示进行 n 次独立的伯努利试验中成功的次数，其中每次试验成功的概率为 p。根据棣莫弗-拉普拉斯定理，当 n 较大时，标准化后的随机变量：

$$Z=\frac{X-np}{\sqrt{np(1-p)}} \tag{20}$$

会趋近于标准正态分布 $N(0,1)$。

例题：一船船在某海区航行，已知每遭遇一次波浪的冲击，纵摇角大于 3° 的概率为 $p=\frac{1}{3}$，若船只遭遇了 90,000 次波浪冲击，问其中有 29,500 ~ 30,500 次纵摇角大于 3° 的概率是多少?

解：

$X\sim b(90000,\frac{1}{3})$。

根据二项分布的分布律，$P(X=k)=C_{90000}^{k}\left(\frac{1}{3}\right)^{k}\left(\frac{2}{3}\right)^{90000-k}$，其中 $k=0,1,\ldots,90000$。

要求 $P(29500\leq X\leq 30500)$：

$$P(29500\leq X\leq 30500)=P\left(\frac{X-90000\times\frac{1}{3}}{\sqrt{90000\times\frac{1}{3}\times\frac{2}{3}}}\leq\frac{30500-90000\times\frac{1}{3}}{\sqrt{90000\times\frac{1}{3}\times\frac{2}{3}}}\right)\tag{21}$$

通过中心极限定理，$X\sim b(90000,\frac{1}{3})$ 近似为：

$$\frac{X-30000}{\sqrt{90000\times\frac{1}{3}\times\frac{2}{3}}}\sim N(0,1)\tag{22}$$

即：

$$\frac{X-30000}{100\sqrt{2}}\sim N(0,1)\tag{23}$$

求概率：

$$P(29500\leq X\leq 30500)=P\left(\frac{29500-30000}{100\sqrt{2}}\leq\frac{X-30000}{100\sqrt{2}}\leq\frac{30500-30000}{100\sqrt{2}}\right)\tag{24}$$

这就转化为：

$$P\left(-\frac{5}{\sqrt{2}}\leq Z\leq\frac{5}{\sqrt{2}}\right)\tag{25}$$

标准正态分布，计算：

$$\Phi\left(\frac{5}{\sqrt{2}}\right)-\Phi\left(-\frac{5}{\sqrt{2}}\right)=2\Phi\left(\frac{5}{\sqrt{2}}\right)-1.\tag{26}$$