第二节　大数定律

一、依概率收敛

二、大数定律

5.2.1 依概率收敛

定义： 设$X_1, X_2, \cdots, X_n, \cdots$是一个随机变量序列，$a$是一个常数，若对于任意给定的正数$\varepsilon$，有

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

或$\lim\limits_{n\to\infty} P\{|X_n - a| \geq \varepsilon\} = 0.$

则称序列$X_1, X_2, \cdots, X_n, \cdots$依概率收敛于$a$，记为

$$X_n \xrightarrow{P} a$$

$X_n \xrightarrow{P} a$：当$n \to \infty$时，X_n落在$(a-\varepsilon, a+\varepsilon)$内的概率越来越大.

而数列$X_n \to a$：$\forall \varepsilon > 0, \exists n_0$，当$n > n_0$时，$|X_n - a| < \varepsilon$

例如：

设n_A表示n次抛硬币试验中正面朝上的次数，则$\frac{n_A}{n} \xrightarrow{P} \frac{1}{2}$，

但$\lim\limits_{n\to\infty} \frac{n_A}{n} \neq \frac{1}{2}$.

考虑极端情况$\frac{n_A}{n} = 1$或$\frac{n_A}{n} = 0$.

性质：若 $X_n \xrightarrow{P} a$，$Y_n \xrightarrow{P} b$.

函数$g(x,y)$在点(a,b)连续，

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

如：当$n \to \infty$时，有 $X_n + Y_n \xrightarrow{P} a + b$.

$$X_n \times Y_n \xrightarrow{P} ab,$$

$$X_n / Y_n \xrightarrow{P} a / b,$$

特别 若 $X_n \xrightarrow{P} a$，函数$f(x)$在点a连续，

则当$n \to \infty$时，$f(X_n) \xrightarrow{P} f(a)$.

5.2.2 大数定律

大量随机现象平均结果的稳定性。

1. 为何能以事件发生的频率作为事件概率的估计？
2. 为何能以算术平均值作为随机变量期望的估计？
3. 为何能以样本均值作为总体期望的估计？

切比雪夫大数定律

辛钦大数定律

伯努利大数定律

切比雪夫定理　设 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立, 分别具有数学期望 $E(X_1), E(X_2), \cdots, E(X_n), \cdots$

及方差 $D(X_1), D(X_2), \cdots, D(X_n), \cdots$

并且方差是一致有上界的, 即存在正数 M, 使得 $D(X_n) \le M$, $n = 1, 2, \cdots$, 则对于任意正数 ε ，恒有

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

证明 $E\left[\frac{1}{n}\sum_{k=1}^{n}X_k\right]=\frac{1}{n}\sum_{k=1}^{n}E(X_k),\qquad D\left[\frac{1}{n}\sum_{k=1}^{n}X_k\right]=\frac{1}{n^2}\sum_{k=1}^{n}D(X_k),$

由**切比雪夫不等式**可得

$$P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right\}\geq 1-\frac{\sum_{k=1}^{n}D(X_k)}{\varepsilon^2 n^2},$$

由方差一致有上界可得

$$\sum_{k=1}^{n}D(X_k)\leq nM,$$

从而 $P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\frac{1}{n}\sum_{k=1}^{n}E(X_k)\right|<\varepsilon\right\}\geq 1-\frac{M}{\varepsilon^2 n},$

在上式中令 $n\to\infty$, 并注意到概率小于等于1, 则

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

切比雪夫定理可叙述成:

设随机变量 $X_1, X_2, \cdots, X_n \cdots$ 相互独立，分别具有期望和方差，且方差一致有上界。则随机变量 $X_1, X_2, \cdots, X_n$ 的算术平均值 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 依概率收敛于它们的数学期望的算术平均值。

推论（切比雪夫定理的特殊情况）

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：$E(X_k) = \mu$，$D(X_k) = \sigma^2 \ (k = 1, 2, \cdots)$，作前 n 个随机变量的算术平均 $\overline{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$， 则对于任意正数 ε 有

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

应用一：

测量值估计：在实际工作中，以大量测量值的算术平均值作为精确值的估计值。

推论的另一种叙述:

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：$E(X_k)=\mu$, $D(X_k)=\sigma^2\ (k=1,2,\cdots)$, 则序列 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于 μ, 即 $\overline{X}\xrightarrow{P}\mu$.

关于推论的说明:

当 n 很大时,随机变量 $X_1, X_2, \cdots, X_n$ 的算术平均 $\frac{1}{n}\sum_{k=1}^{n} X_k$ 接近于数学期望

$$E(X_1) = E(X_2) = \cdots = E(X_k) = \mu,$$

(这个接近是概率意义下的接近)

即在定理条件下, n个随机变量的算术平均, 当n无限增加时, 几乎变成一个常数.

定理二（伯努利(Bernoulli)大数定理） 雅各布•伯努利

设 n_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 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| < \varepsilon\right\} = 1 \text{ 或 } \lim_{n\to\infty} P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} = 0.$$

证法一 引入随机变量

$$X_k = \begin{cases} 0, & \text{若在第} k \text{次试验中} A \text{不发生}, \\ 1, & \text{若在第} k \text{次试验中} A \text{发生}, k = 1,2,\cdots. \end{cases}$$

显然 $\quad n_A = X_1 + X_2 + \cdots + X_n,$

因为 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的,

且 X_k 服从以 p 为参数的 $(0-1)$ 分布,

所以 $E(X_k) = p, \quad D(X_k) = p(1-p), \quad k = 1, 2, \cdots.$

根据推论有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}(X_1 + X_2 + \cdots + X_n) - p\right| < \varepsilon\right\} = 1,$$

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

证法二 显然，$n_A \sim B(n,p)$，已知 $\mathrm{E}(n_A)=np$，$\mathrm{D}(n_A)=np(1-p)$，

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

由切比雪夫不等式 $P\{|X-\mu|<\varepsilon\}\geq 1-\dfrac{\mathrm{D}(X)}{\varepsilon^2}.$

$$1\geq P\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}\geq 1-\frac{D\left(\frac{n_A}{n}\right)}{\varepsilon^2}=1-\frac{p(1-p)}{n\varepsilon^2}$$

在上式令$n\to\infty$取极限即 $\lim\limits_{n\to\infty} P\left\{\left|\dfrac{n_A}{n}-p\right|<\varepsilon\right\}=1$

由伯努利定理的等价形式，

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

当n很大时，事件A在n次独立重复试验中发生的频率与A在一次试验中发生的概率有较大偏差的可能性很小。

关于伯努利定理的说明:

伯努利定理表明事件发生的频率 $\frac{n_A}{n}$ 依概率收敛于事件的概率 p, 它以严格的数学形式表达了频率的稳定性.

应用二：

1. 频率的稳定性：在实际应用中，当试验次数n很大时，可用事件A发生的频率代替A发生的概率。

2. 实际推断原理：概率很小的事件，发生的频率也很小，在一次试验中几乎不可能发生。实际中看做不可能事件。

3. 统计推断时，假设检验以实际推断原理为基础。

定理三（辛钦定理）

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k) = \mu$ $(k = 1,2,\cdots)$，

则对于任意正数 ε，有 $\lim\limits_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right| < \varepsilon\right\} = 1.$

关于辛钦定理的说明:

(1) 不要求方差存在;

(2) 伯努利定理是辛钦定理的特殊情况.

注 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立， 服从同一分布，且具有数学期望 $E(X_k)=\mu \ (k=1,2,\cdots)$,

则前n个随机变量$X_1, X_2, \cdots, X_n$的算术平均值 $\frac{1}{\mathbf{n}}\sum_{i=1}^{n} X_i$

依概率收敛于它们的期望μ.

如果 $E(X_k^l)=\mu_l \ (k=1,2,\cdots)$存在，则

$$\frac{1}{n}\sum_{k=1}^{n} X_k^l \text{依概率收敛于 } \mu_l.$$

应用三：

以上结论说明，样本矩依概率收敛于总体矩。
高阶样本矩依概率收敛于高阶总体矩。
该结论为数理统计中矩估计法的理论基础。

例 设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 独立同分布，且 $E(X_k)=0$, $D(X_k)=\sigma^2, k=1,2,\cdots$, 证明对任意正数 ε 有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \sigma^2\right| < \varepsilon\right\} = 1.$$

解 因为 $X_1, X_2, \cdots, X_n, \cdots$ 是相互独立的，

所以 $X_1^2, X_2^2, \cdots, X_n^2, \cdots$ 也是相互独立的，

由 $E(X_k)=0$, 得 $E(X_k^2) = D(X_k) + [E(X_k)]^2 = \sigma^2$,

由辛钦定理知

对于任意正数 ε, 有 $\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n} X_k^2 - \sigma^2\right| < \varepsilon\right\} = 1.$