第 五 章 大数定律和中心极限定理

第一节 大数定律

1、随机变量序列的收敛

2、概率不等式

3、常见大数定律

4、小结、思考

应 用 背 景

在实际应用中，常遇到如下现象或问题

1 大量测量结果的平均值稳定在一个常数；

2 事件发生的频率稳定于事件发生的概率，“频率
的稳定性”到底是什么意思？实际中有什么用?

3 为什么现实中大量随机变量近似服从正态分布?
在什么情况下可以认为随机变量近似服从正态
分布?

等等，诸如此类的问题。

一、随机变量序列的收敛性

定义5.1.1（依概率收敛）

设$\{X_n\}$是一个随机变量序列，X 是一个随机变量或常数，若对于任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|X_n - X| \geq \varepsilon\} = 0$$

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

则称随机变量序列$\{X_n\}$依概率收敛于X，记为

$$X_n \xrightarrow{P} X \quad \text{或者} \quad \lim_{n\to\infty} X_n = X,\ (P)$$ ■

注：随机变量序列依概率收敛不同于微积分中的数列收敛。

依概率收敛与数列收敛的不同。

1）数列$\{a_n\}$收敛于a是指：当n充分大时，$\forall \varepsilon > 0$

$|a_n - a| < \varepsilon$ 是必然的。

2）随机变量序列$\{X_n\}$依概率收敛于X 是指$\forall \varepsilon > 0$

当n充分大时，事件$\{|X_n - X| < \varepsilon\}$发生的概率很大

（并不是说事件$\{|X_n - X| < \varepsilon\}$一定发生）。

例1. 设随机变量序列 $X_1, X_2, \cdots, X_n \cdots$ 服从如下的分布:

$$P\{X_n = 0\} = 1 - \frac{1}{n} \qquad P\{X_n = 2n\} = \frac{1}{n}$$

则对任意的 $\varepsilon > 0$, 有

$$\lim_{n\to\infty} P\{|X_n| \geq \varepsilon\} = \lim_{n\to\infty} P\{X_n = 2n\} = 0$$

故 $\{X_n\}$ 依概率收敛于 0. 但无论对多大的n, X_n都可能取远离 0 的值 $2n$.

定义5.1.2（依分布收敛）

设随机变量X，X_1，X_2，…的分布函数分别为$F(x), F_1(x), F_2(x), \ldots$,若

$$\lim_{n\to\infty} F_n(x) = F(x)$$

在$F(x)$的每一个连续点上都成立，则称随机变量序列$\{X_k\}, k = 1,2,\ldots$依分布收敛于X。并记为:

$$X_n \xrightarrow{L} X \qquad \blacksquare$$

依分布收敛的意思是指n充分大时：

$$p\{X_n \le x\} \approx p\{X \le x\}$$

二. 概率不等式

1 切比雪夫(*Chebyshev*)不等式

设随机变量 X 的数学期望 $E(X)$ 和方差 $D(X)$都存在, 则对于任意的 $\varepsilon>0$, 有

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

或者

$$P\{|X-E(X)|<\varepsilon\}\geq 1-\frac{D(X)}{\varepsilon^2}$$ ■

例：$X\sim N(\mu,\sigma^2)$，令$\varepsilon=3\sigma$，则有：

$$P\{|X-\mu|<3\sigma\}\geq 1-\frac{\sigma^2}{(3\sigma)^2}=\frac{8}{9}\approx 0.8889$$

切比雪夫不等式的应用

1）估计随机变量分散程度的概率界限，或者说估计随机变量落在某个区间内的概率,但是其估计是很粗略的。

2）理论推导。

例2 设随机变量$X_1, X_2, \cdots, X_{10}$相互独立并且服从相同的分布，已知它们的数学期望等于0，方差等于1，$Y = X_1 + X_2 + \cdots + X_{10}$，请估算概率 $P\{-10<Y<10\}$ 之值。

解 $E(Y) = 0,\ D(Y) = 10,$

$$P\{-10<Y<10\}=P\{|Y|<10\}=P\{|Y-E(Y)|<10\}$$

由切比雪夫不等式，有

$$P\{-10<Y<10\}\geq 1-\frac{10}{10^2}=0.9$$

定义：　大数定律

设$\{X_n\}$是一个随机变量序列，其数学期望都存在，若对于任意的$\varepsilon > 0$，有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i)| < \varepsilon\} = 1$$

则称随机变量序列$\{X_n\}$服从大数定律。 ■

注：1) $\{X_k\}$服从大数定律即是指

$$\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} E(X_i) \xrightarrow{p} 0$$

2) 服从大数定律即是$\{X_k\}$的前n项算术平均值将在概率意义下接近其数学期望$\frac{1}{n}\sum_{i=1}^{n} E(X_i)$

三。常见大数定律

1 切比雪夫大数定律

设$\{X_k\}$是相互独立的随机变量序列，其数学期望和方差都存在，且存在一个常数C，使得

$$D(X_k) < C,\ k = 1,2,\ldots$$

则随机变量序列$\{X_k\}$服从大数定律。 ■

简要说明：

$$E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n}\sum_{i=1}^{n} E(X_i)$$

$$D[\frac{1}{n}\sum_{i=1}^{n} X_i] = \frac{1}{n^2}\sum_{i=1}^{n} D(X_i) \le \frac{nC}{n^2} = \frac{C}{n}$$

由切比雪夫不等式

2 独立同分布大数定律

设$\{X_k\}$是相互独立且同分布的随机变量序列，且$E(X_k)=\mu, D(X_k)=\sigma^2, \ k=1,2,\ldots$

则$\{X_k\}$服从大数定律，即对任意的$\varepsilon>0$,有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k - \mu| < \varepsilon\} = 1 \qquad \blacksquare$$

注：1 此定理为切比雪夫大数定律的一个推论。

2 定理成立时，n个随机变量的算术平均，当n无限增加时将几乎变成一个常数。

3辛钦大数定律

设$\{X_k\}$是相互独立且同分布的随机变量序列，且$E(X_k)=\mu$

则$\{X_k\}$服从大数定律，即对任意的$\varepsilon>0$,有

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^{n} X_k-\mu|<\varepsilon\}=1 \quad \blacksquare$$

注：此定律比独立同分布大数定律适用范围更广。

4 贝努里(*Bernulli*)大数定律

设 $\dfrac{m}{n}$ 是 n 次重复独立试验中事件 A 发生的频率，

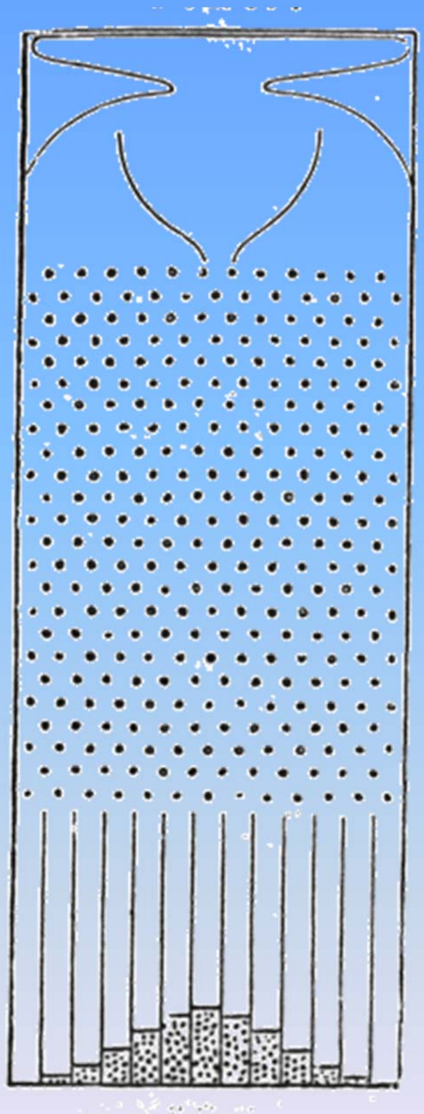
p 是事件 A 在每次试验中发生的概率, 则对于

$\forall \varepsilon > 0$, 有 $$\lim_{n\to\infty} P\{|\frac{m}{n} - p| < \varepsilon\} = 1$$ ■

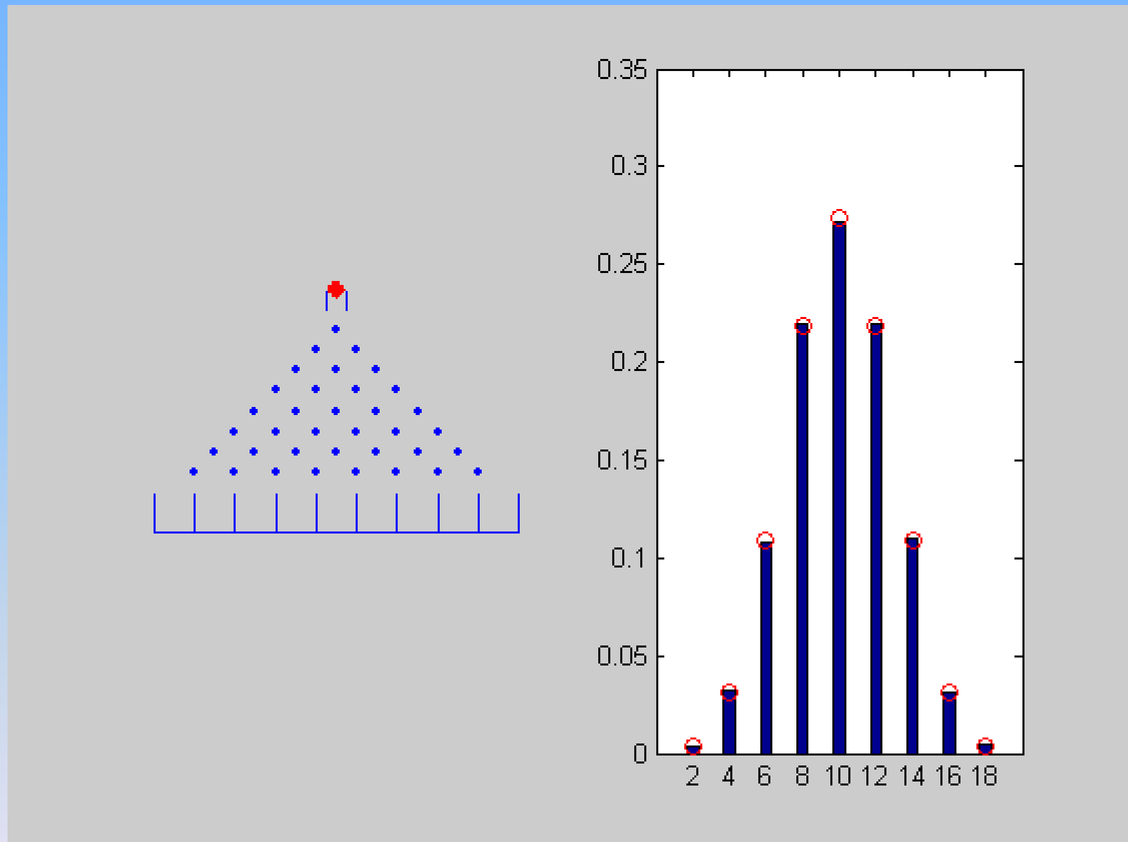
注: 1 此定理为切比雪夫大数定律的一个推论。

2 此定理以严格的数学形式描述了频率的稳定性。

3 由此定理，我们可得小概率事件原理：概率很小的事件，在一次试验中几乎是不可能发生的，从而在实际中可看成不可能事件。



高尔顿板



小　结

1. 切比雪夫不等式主要用来做理论推导和概率估计

2. 常见大数定律关系：

$$
\begin{tikzcd}
 & \boxed{\text{切比雪夫大数定律}} \arrow[d, Rightarrow] \\
\boxed{\text{辛钦大数定律}} \arrow[r, Rightarrow] & \boxed{\text{独立同分布大数定律}} \arrow[d, Rightarrow] \\
 & \boxed{\text{伯努利大数定律}}
\end{tikzcd}
$$