

大数定律

- 引言

大量试验次数下，频率具有稳定性

$$\frac{n_A}{n} \rightarrow p = P(A), (n \rightarrow \infty)$$

依概率收敛

伯努利大数定律

$$\frac{n_A}{n} \rightarrow p = P(A), (n \rightarrow \infty)$$

在 n 重伯努利试验中，令

$$X_k = \begin{cases} 1, & \text{若事件} A \text{在第} k \text{次试验中发生} \\ 0, & \text{若事件} A \text{在第} k \text{次试验中不发生} \end{cases} \quad k = 1, 2, \dots, n$$

$$\rightarrow X_1, X_2, \dots, X_n: i. i. d. \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

$$\rightarrow n_A = \sum_{k=1}^n X_k \quad \frac{n_A}{n} = \frac{1}{n} \sum_{k=1}^n X_k \rightarrow P(A) = p, \quad (n \rightarrow \infty)$$

$$\frac{n_A}{n} = \frac{1}{n} \sum_{k=1}^n X_k \rightarrow p = P(A), (n \rightarrow \infty)$$

$$X_1, X_2, \dots, X_n: i. i. d. \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

$$\longrightarrow E(X_k) = p \longrightarrow E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k) = p$$

$$\longrightarrow \frac{1}{n} \sum_{k=1}^n X_k \rightarrow p = \frac{1}{n} \sum_{k=1}^n E(X_k), \quad (n \rightarrow \infty)$$

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow p = E \left(\frac{1}{n} \sum_{k=1}^n X_k \right), \quad (n \rightarrow \infty)$$

$\{X_n: n = 1, 2, \dots\}$ 是一个随机变量序列

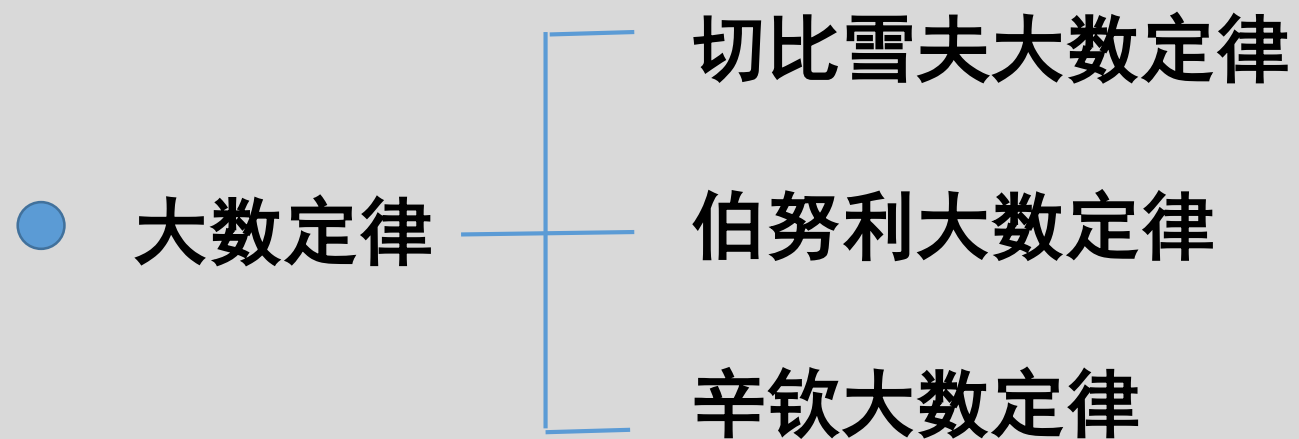
$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow p = \frac{1}{n} \sum_{k=1}^n E(X_k), \quad (n \rightarrow \infty)$$

依概率收敛

大数定律

大数定律

- 依概率收敛的定义



依概率收敛的定义

设 $\{X_n: n = 1, 2, \dots\}$ 为随机变量序列, X 为随机变量. 若对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|X_n| > \varepsilon) = 0$$

则称 X_n 依概率收敛于0, 并记为 $X_n \xrightarrow{P} 0$.

若 $X_n - X \xrightarrow{P} 0$, 则称 X_n 依概率收敛于 X , 并记为 $X_n \xrightarrow{P} X$.

依概率收敛的定义

$X_n \xrightarrow{P} X \iff$ 若对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

\iff 对任意 $\varepsilon > 0$, 当 n 充分大时, 事件

$\{|X_n - X| > \varepsilon\}$ 发生的概率很小.

大数定律

若随机变量序列 $\{X_n: n = 1, 2, \dots\}$ 满足

$$\bar{X} \quad \lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) = 0 \quad E(\bar{X})$$

即
$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \frac{1}{n} \sum_{k=1}^n E(X_k)$$

则称随机变量序列 $\{X_n: n = 1, 2, \dots\}$ 服从大数定律.

切比雪夫大数定律

设 $\{X_n: n = 1, 2, \dots\}$ 为独立随机变量序列. 若存在常数 C , 使得

$D(X_n) \leq C, n = 1, 2, \dots$, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) = 0$$

即
$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \frac{1}{n} \sum_{k=1}^n E(X_k).$$

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) = 0$$

证明：令 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$, 则 $E(\bar{X}) = E \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n} \sum_{k=1}^n E(X_k)$

由于随机变量 X_1, \dots, X_n 独立, 故

$$D(\bar{X}) = D \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n^2} \sum_{k=1}^n D(X_k)$$

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

由切比雪夫不等式，有

$$\begin{aligned}
 & \bar{X} \quad \leftarrow \quad \left[\frac{1}{n} \sum_{k=1}^n X_k \right] \quad \rightarrow \quad E(\bar{X}) \quad \rightarrow \quad D(\bar{X}) \\
 & P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) \leq \frac{\frac{1}{n^2} \sum_{k=1}^n D(X_k)}{\varepsilon^2} \\
 & \left[D(X_n) \leq C, n = 1, 2, \dots, \right] \leq \frac{C}{\varepsilon^2 n} \longrightarrow 0 \quad (n \rightarrow \infty)
 \end{aligned}$$

切比雪夫大数定律

设 $\{X_n: n = 1, 2, \dots\}$ 为独立同分布随机变量序列. $E(X_n) = \mu$,

$D(X_n) = \sigma^2, n = 1, 2, \dots$, 则由切比雪夫大数定律有

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) = 0$$

即 $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu.$

大量试验次数下, 算术平均值稳定性的解释.

若 $\{X_n\}$ 独立同分布, $E(X_n) = \mu$, $D(X_n) = \sigma^2$, 则 $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu$
 $n = 1, 2, \dots,$

在 n 重伯努利试验中, 令

$$X_k = \begin{cases} 1, & \text{若事件} A \text{在第} k \text{次试验中发生} \\ 0, & \text{若事件} A \text{在第} k \text{次试验中不发生} \end{cases} \quad k = 1, \dots, n$$

则 X_1, \dots, X_n 独立同分布, 且 $E(X_k) = p$, $D(X_k) = p(1 - p)$,

记 $n_A = \sum_{k=1}^n X_k$ 由切比雪夫大数定律有 $\frac{n_A}{n} \xrightarrow{P} p$.

伯努利大数定律

设 n_A 为 n 重伯努利试验中事件 A 发生的次数,

又 A 在每次试验中发生的概率为 p ($0 < p < 1$),

则对任意 $\varepsilon > 0$, 有 $\lim_{n \rightarrow \infty} P\left(\left|\frac{n_A}{n} - p\right| < \varepsilon\right) = 1$

大量试验次数下, 频率稳定性的解释.



辛钦大数定律

设 $\{X_n: n = 1, 2, \dots\}$ 为独立同分布随机变量序列. $E(X_n) = \mu$, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left(\left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k) \right| > \varepsilon \right) = 0$$

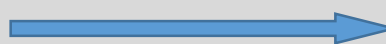
即
$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu.$$

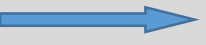
辛钦大数定律

例：设 $\{X_n: n = 1, 2, \dots\}$ 为独立同分布随机变量序列, X_1 服从幂律分布, 其概率密度为

$$f(x) = \begin{cases} \frac{2}{x^3}, & x \geq 1 \\ 0, & x < 1 \end{cases}$$

试证明此随机变量序列服从大数定律, 并求 $\frac{1}{n} \sum_{k=1}^n X_k$ 依概率收敛的极限.

证明: $E(X_1) = \int_1^{\infty} x \frac{2}{x^3} dx = 2$  $\xrightarrow{\text{辛钦大数定律}} \frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} E(X_1) = 2$

$\int_1^{\infty} (x-2)^2 \frac{2}{x^3} dx = \infty$  $D(X_1)$ 不存在, 不满足切比雪夫大数定律

小结

$$X_n \xrightarrow{P} X \iff \text{若对任意 } \varepsilon > 0, \text{ 有}$$
$$\lim_{n \rightarrow \infty} P(|X_n - X| > \varepsilon) = 0$$

大数定律

- 切比雪夫大数定律
- 伯努利大数定律
- 辛钦大数定律