

第五章 大数定律及中心极限定理

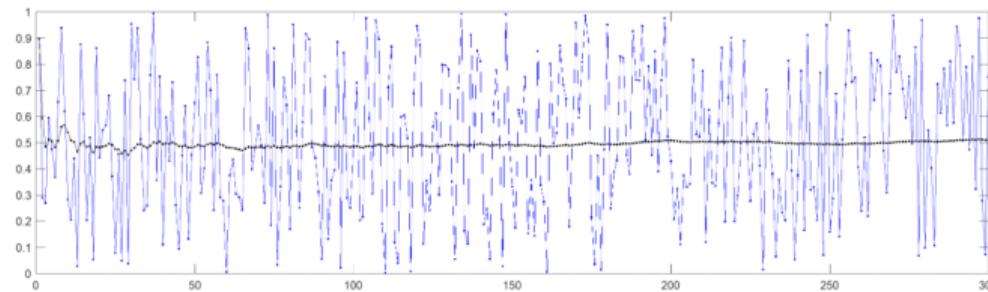
修贤超

<https://xianchaoxiu.github.io>

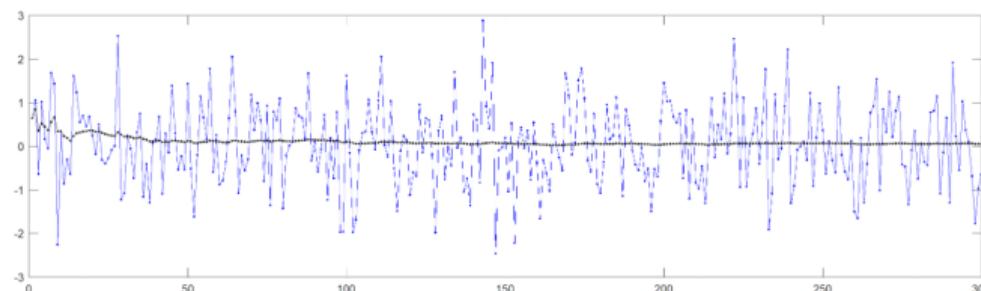
- 5.1 大数定律
- 5.2 中心极限定理

大数定律

■ 蓝点——均匀分布随机变量 黑点——前 n 点平均值



■ 蓝点——正态分布随机变量 黑点——前 n 点平均值



切比雪夫不等式

- 定理：设随机变量 X 具有数学期望 $E(X) = \mu$, 方差 $D(X) = \sigma^2$, 则对于任意正数 ε , 有下列不等式成立

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

证明 根据定义有

$$\begin{aligned} P\{|X - \mu| \geq \varepsilon\} &= \int_{|X-\mu|\geq\varepsilon} f(x)dx \leq \int_{|X-\mu|\geq\varepsilon} \frac{|X - \mu|^2}{\varepsilon^2} dx \\ &\leq \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} |X - \mu|^2 dx = \frac{\sigma^2}{\varepsilon^2} \end{aligned}$$

- 切比雪夫不等式也可写作

$$P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

伯努利大数定律

- 定理：设随机变量 f_A 是 n 次独立重复试验中事件 A 发生的次数, p 是每次试验事件 A 发生的概率, 则对于任意正数 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| < \varepsilon \right\} = 1 \quad \text{或} \quad \lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| \geq \varepsilon \right\} = 0$$

证明 因为 $f_A \sim b(n, p)$, 有 $f_A = X_1 + X_2 + \cdots + X_N$ 且 $E(X_k) = p$
根据切比雪夫不等式, 有

$$\begin{aligned} P \left\{ \left| \frac{f_A}{n} - p \right| < \varepsilon \right\} &= P \{ |f_A - np| < n\varepsilon \} \geq 1 - \frac{np(1-p)}{(n\varepsilon)^2} = 1 - \frac{p(1-p)}{n\varepsilon^2} \\ &\implies \lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| < \varepsilon \right\} = 1 \end{aligned}$$

- n 次伯努利试验, 事件 A 发生的频率随 n 增加, 趋近于概率 p

辛钦大数定律

■ 定理：设 X_1, X_2, \dots 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 $E(X_k) = \mu$ ($k = 1, 2, \dots$). 作前 n 个变量的算术平均 $\frac{1}{n} \sum_{k=1}^n X_k$ ，则对于任意 $\varepsilon > 0$ ，有

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

证明 仅在随机变量的方差 $D(X_k) = \sigma^2$ ($k = 1, 2, \dots$) 存在的条件下证明。
因为

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

由独立性得

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

辛钦大数定律

证明 由切比雪夫不等式得

$$1 \geq \lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} \geq 1 - \frac{\sigma^2/n}{\varepsilon^2}$$

取极限得到

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

- 辛钦大数定律比伯努利大数定律更具有般性
- 定义：设 Y_1, Y_2, \dots 是一个随机变量序列， a 是常数。若对于任意 $\varepsilon > 0$ 有

$$\lim_{n \rightarrow \infty} P \{ |Y_n - a| < \varepsilon \} = 1$$

则称序列 Y_1, Y_2, \dots 依概率收敛于 a ，记作 $Y_n \xrightarrow{P} a$

切比雪夫大数定律

■ 定理: 设 X_1, X_2, \dots 是相互独立的随机变量序列, 且具有数学期望 $E(X_k) = \mu_k$ ($k = 1, 2, \dots$) 及方差 $D(X_k) = \sigma_k^2$ ($k = 1, 2, \dots$). 若存在正数 C , 使 $D(X_k) < C$, 则对于任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n \mu_k \right| < \varepsilon \right\} = 1$$

证明 因为

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

由独立性得

$$D \left(\frac{1}{n} \sum_{k=1}^n X_k \right) = \frac{1}{n^2} \sum_{k=1}^n D(X_k) = \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 \leq \frac{1}{n^2} \sum_{k=1}^n C = \frac{C}{n}$$

切比雪夫大数定律

证明 由切比雪夫不等式得

$$1 \geq \lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n \mu_k \right| < \varepsilon \right\} \geq 1 - \frac{C/n}{\varepsilon^2}$$

取极限得到

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n \mu_k \right| < \varepsilon \right\} = 1$$

- 随机变量序列 X_1, X_2, \dots 的均值 依概率收敛于期望的算数平均值, 即

$$\bar{X} \xrightarrow{P} \bar{\mu}$$

- 切比雪夫大数定律比辛钦大数定律更具有一般性

■ 弱大数定律与强大数定律的区别

□ 弱大数定律: 依概率收敛

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

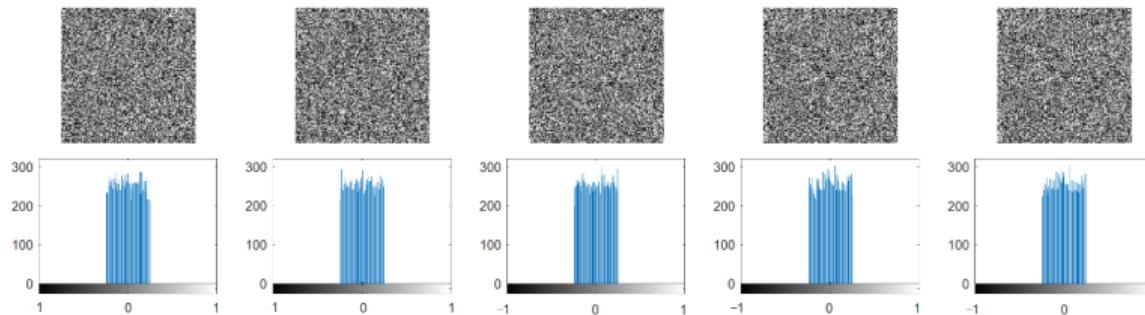
□ 强大数定律: 依概率 1 收敛

$$P \left\{ \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n X_k = \mu \right\} = 1$$

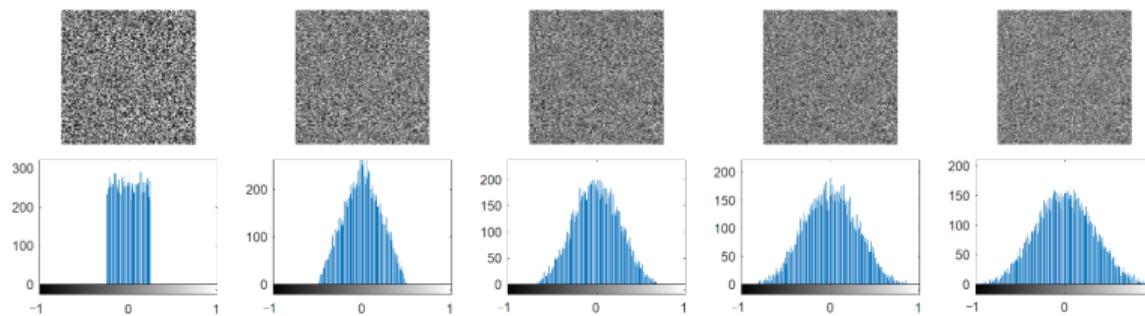
- 5.1 大数定律
- 5.2 中心极限定理

中心极限定理

■ 均匀分布图像噪声



■ 前 n 幅图像噪声的和



独立同分布的中心极限定理

- 定理 1：设 $X_1, X_2, \dots, X_n, \dots$ 相互独立、服从同一分布，且具有数学期望 $E(X_k) = \mu$ ($k = 1, 2, \dots$)，方差 $D(X_k) = \sigma^2$ ($k = 1, 2, \dots$)，则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\begin{aligned}\lim_{n \rightarrow \infty} F_n(x) &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} < x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)\end{aligned}$$

独立同分布的中心极限定理

- 独立同分布随机变量 X_1, X_2, \dots, X_n 之和 $\sum_{k=1}^n X_k$ 的标准化变量, 当 n 充分大时, 近似服从标准正态分布

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \text{ 近似地 } N(0, 1)$$

- 独立同分布随机变量 X_1, X_2, \dots, X_n 的平均值 $\frac{1}{n} \sum_{k=1}^n X_k$ 的标准化变量, 当 n 充分大时, 近似服从正态分布

$$\bar{X} \text{ 近似地 } N(\mu, \sigma^2/n)$$

李雅普诺夫定理

■ 定理 2: 设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 且具有数学期望和方差

$$E(X_k) = \mu_k, D(X_k) = \sigma_k^2 (k = 1, 2, \dots)$$

记 $B_n^2 = \sum_{k=1}^n \sigma_k^2$. 若存在正数 δ , 使得当 $n \rightarrow \infty$ 时,

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E\{|X_k - \mu_k|^{2+\delta}\} \rightarrow 0$$

则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Z_n = \frac{\sum_{k=1}^n X_k - E(\sum_{k=1}^n X_k)}{\sqrt{D(\sum_{k=1}^n X_k)}} = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu}{B_n}$$

的分布函数 $F_n(x)$ 对于任意 x 满足

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu}{B_n} < x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

李雅普诺夫定理

- 独立随机变量 X_1, X_2, \dots, X_n 之和 $\sum_{k=1}^n X_k$ 的标准化变量

$$Z_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu}{B_k}$$

当 n 充分大时, 近似服从正态分布

$$\sum_{k=1}^n X_k \underset{\text{近似地}}{\sim} N\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right) = N\left(\sum_{k=1}^n \mu_k, B_n^2\right)$$

- 与独立同分布的中心极限定理相比, 李雅普诺夫定理更具有般性

棣莫弗-拉普拉斯定理

- 定理 3: 设随机变量 η_n ($n = 1, 2, \dots$) 服从参数 n, p ($0 < p < 1$) 的二项分布, 则对于任意 x , 有

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

证明 引入随机变量

$$X_k = \begin{cases} 1, & \text{第 } k \text{ 次试验 } A \text{ 发生} \\ 0, & \text{第 } k \text{ 次试验 } A \text{ 不发生} \end{cases}$$

于是

$$P\{X_k = 1\} = p, \quad P\{X_k = 0\} = 1 - p$$

$$E(X_k) = p, \quad D(X_k) = p(1-p), \quad \eta_n = \sum_{k=1}^n X_k$$

棣莫弗-拉普拉斯定理

证明 由切比雪夫不等式得

$$\begin{aligned}\lim_{n \rightarrow \infty} P \left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} &= \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - np}{\sqrt{np(1-p)}} \leq x \right\} \\ &= \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)\end{aligned}$$

- 正态分布是二项分布的极限分布
- 当 n 充分大时, 可用正态分布近似计算二项分布的概率

例 1

- 一加法器同时收到 20 个噪声电压 V_k ($k = 1, 2, \dots, 20$), 设他们是互相独立的随机变量, 且都在区间 $(0, 10)$ 服从均匀分布. 记叠加电压 $V = \sum_{k=1}^{20} V_k$, 求 $P\{V > 105\}$ 的近似值

解答 易知

$$\mu = E(V_k) = \frac{10 + 0}{2} = 5$$

$$\sigma^2 = D(V_k) = \frac{(10 - 0)^2}{2}$$

随机变量为

$$Z = \frac{\sum_{k=1}^{20} V_k - 20 \times 5}{\sqrt{100/12}\sqrt{20}} = \frac{V - 20 \times 5}{\sqrt{100/12}\sqrt{20}}$$

例 1

解答 由定理 1, Z 近似服从正态分布 $(0, 1)$, 于是

$$\begin{aligned} P\{V > 105\} &= P\left\{\frac{V - 20 \times 5}{\sqrt{100/12}\sqrt{20}} > \frac{105 - 20 \times 5}{\sqrt{100/12}\sqrt{20}}\right\} \\ &= P\left\{\frac{V - 20 \times 5}{\sqrt{100/12}\sqrt{20}} > 0.387\right\} \\ &= 1 - P\left\{\frac{V - 20 \times 5}{\sqrt{100/12}\sqrt{20}} \leq 0.387\right\} \\ &\approx 1 - \int_{-\infty}^{0.387} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &= 1 - \Phi(0.387) \\ &= 0.348 \end{aligned}$$

例 2

- 一船舶在海上航行，已知每遭受一次海浪冲击，纵摇角大于 3 度的概率为 $p = 1/3$ ，若船遭受了 90000 次波浪冲击，问其中有 29500 30500 次纵摇角大于 3 度的概率是多少

解答 根据定理 3, 有

$$\begin{aligned} P\{29500 < X < 30500\} &= P\left\{\frac{29500 - np}{\sqrt{np(1-p)}} \leq \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{30500 - np}{\sqrt{np(1-p)}}\right\} \\ &= \int_{\frac{29500 - np}{\sqrt{np(1-p)}}}^{\frac{30500 - np}{\sqrt{np(1-p)}}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi\left(\frac{30500 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{29500 - np}{\sqrt{np(1-p)}}\right) \end{aligned}$$

其中 $n = 90000$, $p = 1/3$. 即有

$$P\{29500 < X < 30500\} = \Phi\left(\frac{5\sqrt{2}}{2}\right) - \Phi\left(-\frac{5\sqrt{2}}{2}\right) = 0.995$$

例 3

- 对于一名学生而言, 参加家长会的人数是随机变量, 0、1、2 名家长参会的概率分别是 0.05、0.8、0.15. 学校共有学生 400 名, 各学生参会家长人数相互独立. (1) 求参会家长超过 450 名的概率

解答 设第 k 名学生家长参会人数为随机变量 X_k , 则有

$$\mu = E(X_k) = 0.05 \times 0 + 0.8 \times 1 + 0.15 \times 2 = 1.1$$

$$\begin{aligned}\sigma^2 &= D(X_k) = E(X_k^2) - [E(X_k)]^2 \\ &= 0.05 \times 0^2 + 0.8 \times 1^2 + 0.15 \times 2^2 - 1.1^2 = 0.19\end{aligned}$$

参会家长总人数 $X = \sum_{k=1}^{400} X_k$, 由定理 1, 其标准化随机变量近似服从正态分布 $(0, 1)$, 于是

$$\begin{aligned}P\{V > 450\} &= P\left\{Z > \frac{450 - 400 \times 1.1}{\sqrt{400} \sqrt{0.19}}\right\} = P\{Z > 1.147\} \\ &= 1 - \Phi(1.147) = 0.1251\end{aligned}$$

例 3

- 对于一名学生而言, 参加家长会的人数是随机变量, 0、1、2 名家长参会的概率分别是 0.05、0.8、0.15. 学校共有学生 400 名, 各学生参会家长人数相互独立. (2) 求一名家长参会学生人数不多于 340 的概率

解答 设有 1 名学生家长参会人数为随机变量 Y , 则有 $Y \sim b(400, 0.8)$. 由定理 3 知

$$\begin{aligned} P\{Y \leq 450\} &= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \leq \frac{349 - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}}\right\} \\ &= P\left\{\frac{Y - 400 \times 0.8}{\sqrt{400 \times 0.8 \times 0.2}} \leq 2.5\right\} \\ &= \Phi(2.5) \\ &= 0.9938 \end{aligned}$$

■ 考试内容

- 切比雪夫 (Chebyshev) 不等式、切比雪夫大数定律、伯努利 (Bernoulli) 大数定律、辛钦 (Khinchine) 大数定律、棣莫弗—拉普拉斯 (DeMoivre — Laplace) 定理、列维—林德伯格 (Levy-Lindberg) 定理

■ 考试要求

- 了解切比雪夫不等式
- 了解切比雪夫大数定律、伯努利大数定律和辛钦大数定律 (独立同分布随机变量序列的大数定律)
- 了解棣莫弗-拉普拉斯定理 (二项分布以正态分布为极限分布) 和列维-林德伯格定理 (独立同分布随机变量序列的中心极限定理)

Q&A

Thank you!

感谢您的聆听和反馈