

概率论 & 数理统计

2025-2026 秋冬

黄炜
(hwmath2003@zju.edu.cn)

浙江大学数学科学学院

Chapter 5 Laws of Large Numbers and Central Limit Theorems

关键词 (Keywords):

- 依概率收敛 (Convergence in Probability)
- Markov 不等式和 Chebyshev 不等式 (Markov Inequality and Chebyshev Inequality)
- 大数定律 (Laws of Large Numbers)
- 中心极限定理 (Central Limit Theorems)

(后两者为极限定理的其中两种)

§5.1 大数定律 (Laws of Large Numbers)

内容:

设 $X_1, X_2, \dots, X_n, \dots$ 是一列随机变量, 则在一定的条件下随机变量序列

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \text{ 收敛到 } \mu \text{ (当 } n \rightarrow +\infty \text{ 时).}$$

问题:

- ① 随机变量序列 Y_n 收敛到 μ 的含义?
- ② μ 是什么?
- ③ 一定的条件是什么?

依概率收敛 (Convergence in Probability)

Definition 1 (依概率收敛)

设 $Y_1, Y_2, \dots, Y_n, \dots$ 为一列随机变量, Y 为随机变量, 若对于 $\forall \varepsilon > 0$, 均有

$$\lim_{n \rightarrow +\infty} P\{|Y_n - Y| \geq \varepsilon\} = 0$$

成立, 则称随机变量序列 $\{Y_n, n \geq 1\}$ 依概率收敛于 Y (Y_n converge to Y in probability), 记为 $Y_n \xrightarrow{P} Y$, 当 $n \rightarrow +\infty$.

特别地, 当 $Y = c$ 为一常数时, 称 $\{Y_n, n \geq 1\}$ 依概率收敛于常数 c .

注意: 这种收敛性是在概率意义下的一种收敛, 而不是在数学意义上的一般收敛.

Example 1

设 $Y_n \sim N\left(0, \frac{1}{n}\right)$, $n = 1, 2, \dots$, 证明: 当 $n \rightarrow +\infty$ 时, $Y_n \xrightarrow{\text{P}} 0$.

Proof

对任意 $\varepsilon > 0$,

$$\begin{aligned} \mathsf{P}(|Y_n - 0| \geq \varepsilon) &= \mathsf{P}(Y_n \geq \varepsilon) + \mathsf{P}(Y_n \leq -\varepsilon) \\ &= 1 - \Phi\left(\frac{\varepsilon - 0}{\sqrt{1/n}}\right) + \Phi\left(\frac{-\varepsilon - 0}{\sqrt{1/n}}\right) \\ &= 2[1 - \Phi(\varepsilon\sqrt{n})] \rightarrow 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.} \end{aligned}$$

所以当 $n \rightarrow +\infty$ 时, $Y_n \xrightarrow{\text{P}} 0$.

依概率收敛的性质

若当 $n \rightarrow +\infty$ 时, $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, 且函数 $g(x, y)$ 在点 (a, b) 处连续, 则

$$g(X_n, Y_n) \xrightarrow{P} g(a, b), \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

如: 当 $n \rightarrow +\infty$ 时, $X_n \xrightarrow{P} a$, $Y_n \xrightarrow{P} b$, 则

$$X_n + Y_n \xrightarrow{P} a + b, \quad \text{当 } n \rightarrow +\infty \text{ 时,}$$

$$X_n \cdot Y_n \xrightarrow{P} a \cdot b, \quad \text{当 } n \rightarrow +\infty \text{ 时,}$$

$$X_n / Y_n \xrightarrow{P} a/b (b \neq 0), \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

特别地, 若当 $n \rightarrow +\infty$ 时, $X_n \xrightarrow{P} a$, 且 $f(x)$ 在点 a 处连续, 则 $f(X_n) \xrightarrow{P} f(a)$, 当 $n \rightarrow +\infty$ 时.

Markov 不等式 (Markov Inequality)

Theorem 1 (Markov 不等式)

设随机变量 Y 的 $k(k \geq 1)$ 阶矩存在, 即 $E(Y^k)$ 存在, 则对于任意 $\varepsilon > 0$, 有

$$P\{|Y| \geq \varepsilon\} \leq \frac{E(|Y|^k)}{\varepsilon^k}$$

成立.

定理的等价形式为 $P\{|Y| < \varepsilon\} \geq 1 - \frac{E(|Y|^k)}{\varepsilon^k}$.

特别地, 当 Y 为取非负值的随机变量时, 则有 $P\{Y \geq \varepsilon\} \leq \frac{E(Y^k)}{\varepsilon^k}$.

Proof

对于 $\varepsilon > 0$, 令 $Z = \begin{cases} \varepsilon, & \text{当 } |Y| \geq \varepsilon \text{ 时}, \\ 0, & \text{当 } |Y| < \varepsilon \text{ 时}, \end{cases}$ 则 $0 \leq Z \leq |Y|$.

于是对任意的 $k \geq 1$, 当 $E(|Y|^k)$ 存在时, $E(Z^k)$ 也存在, 且 $E(Z^k) \leq E(|Y|^k)$.
而根据 Z 的定义

$$E(Z^k) = \varepsilon^k P\{|Y| \geq \varepsilon\},$$

因此

$$P\{|Y| \geq \varepsilon\} = \frac{E(Z^k)}{\varepsilon^k} \leq \frac{E(|Y|^k)}{\varepsilon^k}.$$

切比雪夫不等式 (Chebyshev Inequality)

Theorem 2 (Chebyshev 不等式)

设随机变量 X 具有数学期望 $E(X) = \mu$, 方差 $\text{Var}(X) = \sigma^2$, 则对于任意 $\varepsilon > 0$, 都有

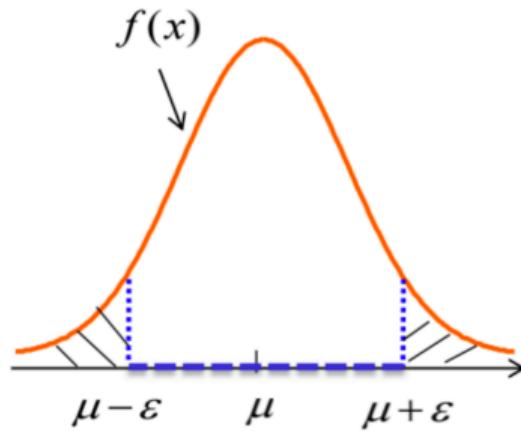
$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}.$$

定理的等价形式为 $P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$.

Proof

在 Markov 不等式中取 $Y = X - \mu, k = 2$ 即可得到 Chebyshev 不等式.
或类似 Markov 不等式的证明也可直接证得.

Chebyshev 不等式的直观意义



Example 2

设 $E(X) = \mu$, $\text{Var}(X) = \sigma^2$, 取 $\varepsilon = K\sqrt{\text{Var}(X)} = K\sigma$, 则有

$$P(|X - \mu| \geq K\sigma) \leq \frac{\text{Var}(X)}{K^2\sigma^2} = \frac{1}{K^2}.$$

若取 $K = 3$, 则有

$$P(|X - \mu| \geq 3\sigma) = P(X \notin (\mu - 3\sigma, \mu + 3\sigma)) \leq \frac{1}{9},$$

而当 $X \sim N(\mu, \sigma^2)$ 时,

$$P(|X - \mu| \geq 3\sigma) = 2(1 - \Phi(3)) = 0.0027 < \frac{1}{9}.$$

Chebyshev 不等式的适用面广泛, 但结果较为粗糙.

Example 3

在 n 重 Bernoulli 试验中, 若已知每次试验事件 A 出现的概率为 0.75, 试利用 Chebyshev 不等式,

- ① 若 $n = 7500$, 估计 A 出现的频率在 0.74 至 0.76 之间的概率至少有多大?
- ② 估计 n , 使得 A 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90.

Solution

设在 n 重 Bernoulli 试验中, 事件 A 出现的次数为 X_n , 则 $X_n \sim B(n, 0.75)$, 于是

$$\mathbb{E}(X_n) = np = 0.75n, \quad \text{Var}(X_n) = np(1-p) = 0.1875n.$$

又 $f_n(A) = X_n/n$.

Solution (Cont.)

故由 Chebyshev 不等式, 可得

$$\mathsf{P}\left\{0.74 < \frac{X_n}{n} < 0.76\right\} = \mathsf{P}\{|X_n - 0.75n| < 0.01n\} \geq 1 - \frac{0.1875n}{(0.01n)^2} = 1 - \frac{1875}{n}.$$

- ① 已知 $n = 7500$, 那么此概率至少为 $1 - \frac{1875}{7500} = 0.75$.
- ② 若要求此概率不小于 0.90, 现 $1 - \frac{1875}{n} \geq 0.90 \Rightarrow n \text{需满足 } n \geq 18750$.

Example 4

某天文机构想测量宇宙中两颗行星的距离, 进行了 n 次独立的观测, 测量值分别为 X_i (光年), $i = 1, 2, \dots, n$. 若 $E(X_i) = \mu$ (为两颗行星的真实距离, 未知), $\text{Var}(X_i) = 5$. 现取这 n 次观测的平均作为实际距离 μ 的估计.

- ① 若 $n = 100$, 则估计值与实际值间的误差在 ± 0.5 光年之内的概率至少有多大?
- ② 若要以不低于 95% 的把握控制估计值与实际值间的误差在 ± 0.5 光年内, 至少要观测多少次?

Solution

由于对 $i = 1, 2, \dots, n$, 都有 $E(X_i) = \mu$, $\text{Var}(X_i) = 5$, 且诸 X_i 相互独立, 故

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \mu, \quad \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{5}{n}.$$

Solution (Cont.)

由 Chebyshev 不等式,

$$\mathbb{P}\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < 0.5\right\} \geq 1 - \frac{5/n}{0.5^2} = 1 - \frac{20}{n}.$$

- ① 已知 $n = 100$, 因此此概率至少为 $1 - \frac{20}{100} = 0.8$.
- ② 若要求此概率不小于 0.95, 现 $1 - \frac{20}{n} \geq 0.95 \Rightarrow n \text{ 需满足 } n \geq 400$.

Example 5

设 X_1, \dots, X_n 是随机变量序列, $E(X_n) = 0$, $\text{Var}(X_n) = \frac{1}{n}$, 证明 $X_n \xrightarrow{\text{P}} 0$, 当 $n \rightarrow +\infty$.

Proof

利用 Chebyshev 不等式, 对于任意的 $\forall \varepsilon > 0$,

$$0 \leq P(|X_n - 0| \geq \varepsilon) \leq \frac{\text{Var}(X_n)}{\varepsilon^2} = \frac{1}{n\varepsilon^2} \rightarrow 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

因此 $\lim_{n \rightarrow +\infty} P(|X_n - 0| \geq \varepsilon) = 0$, 即 $X_n \xrightarrow{\text{P}} 0$, 当 $n \rightarrow +\infty$ 时.

几种大数定律

Definition 2 (大数定律)

设 $\{X_i, i \geq 1\}$ 为一随机变量序列, 若存在常数序列 $\{c_n, n \geq 1\}$, 使得对 $\forall \varepsilon > 0$, 均有

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - c_n\right| \geq \varepsilon\right\} = 0 \text{ 或 } \lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - c_n\right| < \varepsilon\right\} = 1$$

成立, 即有 $\frac{1}{n} \sum_{i=1}^n X_i - c_n \xrightarrow{P} 0$, 当 $n \rightarrow +\infty$ 时, 则称随机变量序列 $\{X_i, i \geq 1\}$ 服从
(弱) 大数定律 (weak law of large numbers).

特别地, 若 c_n 与 n 无关, 则可以写为 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} c$, 当 $n \rightarrow +\infty$ 时.

Theorem 3 (贝努里 (Bernoulli) 大数定律)

设 n_A 为 n 重 Bernoulli 试验中事件 A 发生的次数，并记事件 A 在每次试验中发生的概率为 p ，则对 $\forall \varepsilon > 0$ ，有

$$\lim_{n \rightarrow +\infty} P\left\{ \left| \frac{n_A}{n} - p \right| \geq \varepsilon \right\} = 0$$

成立。即，事件的频率 $\frac{n_A}{n} \xrightarrow{P}$ 事件概率 p ，当 $n \rightarrow \infty$ 。

Proof

引入随机变量 $X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中 } A \text{ 发生,} \\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不发生,} \end{cases} \quad i = 1, 2, \dots, n$ ，则 $X_i \sim B(1, p)$

且相互独立， $E(X_i) = p$, $\text{Var}(X_i) = p(1-p)$, $i = 1, \dots, n$ ，且 $n_A = \sum_{i=1}^n X_i$.

Proof (Cont.)

故 $\frac{n_A}{n} = \frac{1}{n} \sum_{i=1}^n X_i$, 且 $E\left(\frac{n_A}{n}\right) = p$, $\text{Var}\left(\frac{n_A}{n}\right) = \frac{p(1-p)}{n}$.

由 Chebyshev 不等式, 可知

$$0 \leq P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} \leq \frac{\text{Var}(n_A/n)}{\varepsilon^2} = \frac{p(1-p)}{n\varepsilon^2} \rightarrow 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

因此有 $\lim_{n \rightarrow +\infty} P\left\{\left|\frac{n_A}{n} - p\right| \geq \varepsilon\right\} = 0$.

Bernoulli 大数定律的重要意义

- Bernoulli 大数定律证明了在大量重复独立试验中事件发生的频率的稳定性，正因为这种稳定性，概率的概念才有客观意义.
- Bernoulli 大数定律还提供了通过试验来确定事件概率的方法，既然当 n 充分大时，频率 n_A/n 与概率 p 有较大偏差的可能性很小，因此可以通过做试验确定某事件发生的频率并把它作为相应的概率估计，这是一种参数估计方法，该方法的重要理论基础之一就是大数定律.

Theorem 4 (辛钦 (Khinchin) 大数定律)

设 $\{X_i, i \geq 1\}$ 为独立同分布的随机变量序列, 且其期望存在, 记为 μ , 则对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| \geq \varepsilon\right\} = 0$$

成立, 即 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \mu$, 当 $n \rightarrow +\infty$ 时.

即随机变量序列 $\{X_i, i \geq 1\}$ 服从大数定律.

Khinchin 大数定律的证明需要用到特征函数 (characteristic function), 感兴趣的同学可以自行了解, 此处从略.

Corollary 1

设 $\{X_i, i \geq 1\}$ 为独立同分布的随机变量序列, 若 $h(x)$ 为一连续函数, 且 $E(|h(X_1)|) < +\infty$, 则对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left\{ \left| \frac{1}{n} \sum_{i=1}^n h(X_i) - a \right| \geq \varepsilon \right\} = 0$$

成立, 其中 $a = E(h(X_1))$, 即 $\frac{1}{n} \sum_{i=1}^n h(X_i) \xrightarrow{P} a = E(h(X_1))$, 当 $n \rightarrow +\infty$ 时.
 即随机变量序列 $\{h(X_i), i \geq 1\}$ 也服从大数定律.

Khinchin 大数定律及其推论很常用, 值得理解并熟记.

Example 6

设随机变量 X_1, \dots, X_n, \dots 相互独立同分布, $X_1 \sim U(-1, 1)$. 请问: (1) $\frac{1}{n} \sum_{k=1}^n X_k$,
 (2) $\frac{1}{n} \sum_{k=1}^n |X_k|$, (3) $\frac{1}{n} \sum_{k=1}^n X_k^2$ 分别依概率收敛吗? 如果依概率收敛, 分别收敛于什么?

Solution

注意到 X_1, \dots, X_n, \dots 相互独立同分布, $E(X_i)$ 存在;

$|X_1|, \dots, |X_n|, \dots$ 相互独立同分布, $E(|X_i|)$ 存在;

$X_1^2, \dots, X_n^2, \dots$ 相互独立同分布, $E(X_i^2)$ 存在.

因此利用辛钦大数定律, 可知 $\frac{1}{n} \sum_{k=1}^n X_k$, $\frac{1}{n} \sum_{k=1}^n |X_k|$, $\frac{1}{n} \sum_{k=1}^n X_k^2$ 均依概率收敛.

Solution (Cont.)

计算得

$$\mathbb{E}(X_1) = 0, \quad \mathbb{E}(|X_1|) = \int_{-1}^1 \frac{|x|}{2} dx = \frac{1}{2}, \quad \mathbb{E}(X_1^2) = \int_{-1}^1 \frac{x^2}{2} dx = \frac{1}{3}.$$

因此

$$\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{\text{P}} 0, \quad \frac{1}{n} \sum_{k=1}^n |X_k| \xrightarrow{\text{P}} \frac{1}{2}, \quad \frac{1}{n} \sum_{k=1}^n X_k^2 \xrightarrow{\text{P}} \frac{1}{3},$$

当 $n \rightarrow +\infty$ 时.

Example 7

设随机变量 X_1, \dots, X_n, \dots 相互独立同分布, $X_1 \sim U(0, 1)$, 请问: 当 $n \rightarrow +\infty$ 时, $\sqrt[n]{X_1 X_2 \cdots X_n}$ 依概率收敛吗? 如果依概率收敛, 收敛于什么?

Solution

记 $Y_n = \sqrt[n]{X_1 X_2 \cdots X_n}$, 令

$$Z_n = \ln Y_n = \frac{1}{n}(\ln X_1 + \ln X_2 + \cdots + \ln X_n).$$

由 X_1, \dots, X_n, \dots 相互独立同分布, 可知 $\ln X_1, \dots, \ln X_n, \dots$ 相互独立同分布, 且 $E(\ln X_1) = \int_0^1 \ln x dx = -1$, 那么利用辛钦大数定律知 $Z_n \xrightarrow{\text{P}} -1$, 当 $n \rightarrow +\infty$.

进一步利用依概率收敛的性质, 可知 $Y_n = e^{Z_n} \xrightarrow{\text{P}} e^{-1}$, 当 $n \rightarrow +\infty$.

Theorem 5 (*Markov 大数定律, 仅需了解)

设 $\{X_i, i \geq 1\}$ 为一随机变量序列, 若对 $\forall i \geq 1$, X_i 的方差都存在, 并且

$$\lim_{n \rightarrow +\infty} \frac{1}{n^2} \operatorname{Var}\left(\sum_{i=1}^n X_i\right) = 0,$$

则对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| \geq \varepsilon\right\} = 0$$

成立, 即随机变量序列 $\{X_i, i \geq 1\}$ 服从大数定律.

Proof

记 $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$, 则 $E(Y_n) = \frac{1}{n} \sum_{i=1}^n E(X_i)$, $\text{Var}(Y_n) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right)$.

对 Y_n 应用 Chebyshev 不等式, 并结合定理条件, 可得

$$0 \leq P\{|Y_n - E(Y_n)| \geq \varepsilon\} \leq \frac{\text{Var}(Y_n)}{\varepsilon^2} = \frac{1}{n^2 \varepsilon^2} \text{Var}\left(\sum_{i=1}^n X_i\right) \rightarrow 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

即

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| \geq \varepsilon\right\} = 0.$$

Theorem 6 (*Chebyshev 大数定律, 仅需了解)

设 $\{X_i, i \geq 1\}$ 为相互独立的随机变量序列, 若存在常数 C , 使得 $\text{Var}(X_i) \leq C$, $i = 1, 2, \dots$, 即所有的 X_i 的方差有共同的上界, 则对 $\forall \varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n E(X_i)\right| \geq \varepsilon\right\} = 0$$

成立, 即随机变量序列 $\{X_i, i \geq 1\}$ 服从大数定律.

Proof

由于 $\frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \leq \frac{nC}{n^2} = \frac{C}{n} \rightarrow 0$, 当 $n \rightarrow +\infty$ 时.

因此由 Markov 大数定律知结论成立.

Corollary 2 (*Chebyshev 大数定律的推论, 仅需了解)

设 $\{X_i, i \geq 1\}$ 为相互独立的随机变量序列, 若它们的方差存在并相同 (记为 σ^2), 则随机变量序列 $\{X_i, i \geq 1\}$ 也服从大数定律.

Proof

记 $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$, 则 $E(Y_n) = \frac{1}{n} \sum_{i=1}^n E(X_i)$, $\text{Var}(Y_n) = \frac{1}{n^2} \text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{\sigma^2}{n}$.

对 Y_n 应用 Chebyshev 不等式, 并结合定理条件, 可得

$$0 \leq P\{|Y_n - E(Y_n)| \geq \varepsilon\} \leq \frac{\text{Var}(Y_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \rightarrow 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

因此随机变量序列 $\{X_i, i \geq 1\}$ 也服从大数定律.

§5.1 Laws of Large Numbers

下面这个结论较为常用.

特别地, $\{X_i, i \geq 1\}$ 为相互独立的随机变量序列, 具有相同的期望 μ , 相同的方差 σ^2 , 那么

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{P}} \mu, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

Example 8

设随机变量 X_1, \dots, X_n, \dots 相互独立, 且它们的分布律为

$$\mathsf{P}(X_i = \sqrt{i}) = \mathsf{P}(X_i = -\sqrt{i}) = \frac{1}{2i}, \quad \mathsf{P}(X_i = 0) = 1 - \frac{1}{i}, \quad i = 1, 2, \dots$$

试判断 $\{X_i, i \geq 1\}$ 是否服从大数定律?

Solution

由于对任意的 $i \geq 1$, 有

$$\mathbb{E}(X_i) = 0, \quad \text{Var}(X_i) = \mathbb{E}(X_i^2) = 0 + (\sqrt{i})^2 \cdot \frac{1}{2i} + (-\sqrt{i})^2 \cdot \frac{1}{2i} = 1,$$

所以 $\{X_i, i \geq 1\}$ 相互独立, 期望存在并且相同, 方差存在并且相同, 由 Chebyshev 大数定律的推论知满足大数定律, 且

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{P}} 0, \quad \text{当 } n \rightarrow +\infty \text{ 时.}$$

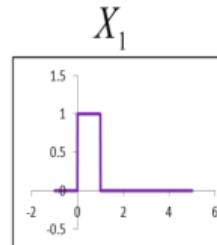
§5.2 Central Limit Theorem

背景: 有许多随机变量, 它们是由大量的相互独立的随机变量的综合影响所形成的, 而其中每个个别的因素作用都很小, 这种随机变量往往服从或近似服从正态分布, 或者说它的极限分布是正态分布.

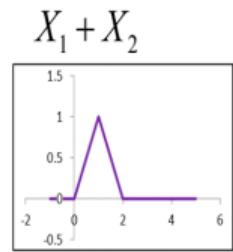
中心极限定理 (Central Limit Theorem, CLT)

正是从数学上论证了这一现象, 它在长达两个世纪的时期内曾是概率论研究的中心课题.

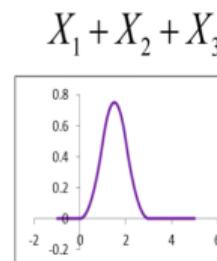
n 个独立同分布的均匀分布 $U(0, 1)$ 随机变量之和:



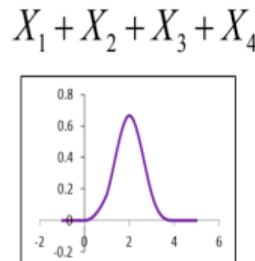
$$X_1$$



$$X_1 + X_2$$



$$X_1 + X_2 + X_3$$



$$X_1 + X_2 + X_3 + X_4$$

Theorem 7 (Lindeberg-Lévy 中心极限定理, 亦称为独立同分布的中心极限定理)

设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立且同分布, 满足 $E(X_i) = \mu, \text{Var}(X_i) = \sigma^2, i = 1, 2, \dots$, 则对于充分大的 n 有

$$\sum_{i=1}^n X_i \xrightarrow{\text{近似}} N(n\mu, n\sigma^2).$$

此时

$$P\left(a < \sum_{i=1}^n X_i \leq b\right) \approx \Phi\left(\frac{b - n\mu}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{a - n\mu}{\sqrt{n}\sigma}\right).$$

类似地, 有 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ 的近似分布是 $N\left(\mu, \frac{\sigma^2}{n}\right)$.

§5.2 Central Limit Theorem

注意到：当随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立同分布，满足 $E(X_i) = \mu$, $\text{Var}(X_i) = \sigma^2$, $i = 1, 2, \dots$ 时，

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) = n\mu, \quad \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = n\sigma^2.$$

$$E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = \frac{1}{n} \cdot n\mu = \mu,$$

$$\text{Var}(\bar{X}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{\sigma^2}{n}.$$

据 CLT, 当 n 充分大时, 有 $\sum_{i=1}^n X_i \xrightarrow{\text{近似}} N(n\mu, n\sigma^2)$, $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{\text{近似}} N\left(\mu, \frac{\sigma^2}{n}\right)$

故 CLT 仅仅是分布类型上的一种近似. (“万物归一”)

Corollary 3 (De Moivre-Laplace 中心极限定理)

设 n_A 为在 n 重 Bernoulli 试验中事件 A 发生的次数, $P(A) = p(0 < p < 1)$, 则对任何实数 x , 有

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

Proof

引入随机变量 $X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验时 } A \text{ 发生,} \\ 0, & \text{第 } i \text{ 次试验时 } A \text{ 未发生,} \end{cases}$ 则 $X_1, X_2, \dots, X_n, \dots$ 相互独立同分布, $X_i \sim B(1, p)$. 由于 $n_A = X_1 + X_2 + \dots + X_n$, 由 Lindeberg-Lévy 中心极限定理可知 $n_A \xrightarrow{\text{近似}} N(np, np(1-p))$.

说明: 当 n 很大时, 二项分布可以用正态分布去近似. (期望与方差不变)

Example 9

某宴会上提供一瓶 6 升 (L) 的大瓶法国红酒，假定与会者每次所倒红酒的重量服从同一分布，期望值为 100 毫升 (mL)，标准差为 32 毫升 (mL)。若每次所倒红酒都是相互独立的，试问：倒了 55 次后该瓶红酒仍有剩余的概率约为多少？

Solution

设 X_i 为第 i 次所倒的红酒重量 (单位: mL)，则 X_i 相互独立且分布相同，且 $E(X_i) = 100$, $\text{Var}(X_i) = 32^2$, $i = 1, 2, \dots, 55$.

由 Lindeberg-Lévy 中心极限定理，

$$\frac{\sum_{i=1}^{55} X_i - 55 \times 100}{32\sqrt{55}} \underset{\text{近似}}{\sim} N(0, 1).$$

Solution (Cont.)

所以

$$\begin{aligned}
 & P\{\text{倒了 } 55 \text{ 次后该瓶红酒仍有剩余}\} \\
 &= P\left\{ \sum_{i=1}^{55} X_i < 6000 \right\} \\
 &= P\left\{ \frac{\sum_{i=1}^{55} X_i - 55 \times 100}{32\sqrt{55}} < \frac{6000 - 55 \times 100}{32\sqrt{55}} \right\} \\
 &\approx \Phi\left(\frac{6000 - 55 \times 100}{32\sqrt{55}}\right) = \Phi(2.11) = 0.9826.
 \end{aligned}$$

Example 10

某校 1500 名学生选修“概率论与数理统计”课程，共有 10 名教师主讲此课，假定每位学生可以随意地选择一位教师（即选择任意一位教师的可能性均为 $1/10$ ），而且学生之间的选择是相互独立的。问：每位教师的上课教室应该设有多少座位才能保证因没有座位而使学生离开的概率小于 5%。

Solution

由于每位学生可以随意地选择一位老师，因此我们只需要考虑某个老师甲的上课教室的座位即可。引入随机变量

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 个学生选择教师甲,} \\ 0, & \text{其他,} \end{cases} \quad i = 1, 2, \dots, 1500,$$

则 X_i 独立同分布， $X_i \sim B(1, 1/10)$ 。

Solution (Cont.)

记 $Y = \sum_{i=1}^{1500} X_i$, 则 $Y \sim B(1500, 1/10)$, 由 De Moivre-Laplace 中心极限定理,

$$\frac{Y - 1500 \times 1/10}{\sqrt{1500 \times 1/10 \times 9/10}} \text{ 近似 } N(0, 1).$$

设教室需要设 a 个座位, 则由题意, a 需要满足

$$\Phi(1.645) = 95\% \leq P\{Y \leq a\} \approx \Phi\left(\frac{a - 1500 \times 1/10}{\sqrt{1500 \times 1/10 \times 9/10}}\right) = \Phi\left(\frac{a - 150}{\sqrt{135}}\right),$$

即 $\frac{a - 150}{\sqrt{135}} \geq 1.645 \Rightarrow a \geq 169.11$. 故每位教师的上课教室应该至少设 170 个座位才能保证因没有座位而使学生离开的概率小于 5%.

Example 11

某保险公司的老年人寿保险有 1 万人参加 (这些人的健康状况相互独立), 每人每年交 200 元, 若老人在该年内死亡, 公司付给受益人 1 万元. 设老年人死亡率为 0.017, 试求保险公司在一年内这项保险亏本的概率约为多少?

Solution

设 X 为一年中投保老人的死亡数, 则 $X \sim B(10000, 0.017)$. 现求 $P(X > 200)$. 由 De Moivre-Laplace 中心极限定理,

$$\frac{X - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}} \stackrel{\text{近似}}{\sim} N(0, 1).$$

故保险公司亏本的概率为

$$P(X > 200) \approx 1 - \Phi\left(\frac{200 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\right) = 1 - \Phi(2.321) = 0.01.$$

Example 10 (Cont.)

求保险公司盈利 10 万元以上的概率.

Solution (Cont.)

类似地, 保险公司盈利 10 万元以上的概率

$$\begin{aligned}& \mathsf{P}(10000 \times 200 - 10000X > 100000) \\&= \mathsf{P}(X < 190) = \mathsf{P}(X \leq 189) = \mathsf{P}(0 \leq X \leq 189) \\&\approx \Phi\left(\frac{189 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\right) - \Phi\left(\frac{0 - 10000 \times 0.017}{\sqrt{10000 \times 0.017 \times 0.983}}\right) \\&= \Phi(1.47) - \Phi(-13.15) = 0.9292\end{aligned}$$

Example 12

一颗骰子投掷了 11520 次，出现“1 点”的次数超过了 2160 次，由此可否推断该骰子是不均匀的呢？

Solution

记 n_A 为骰子投掷了 11520 次出现“1 点”的次数。假设骰子是均匀的，则 $n_A \sim B(11520, 1/6)$ 。那么

$$\mathbb{P}\{n_A > 2160\} = \sum_{k=2161}^{11520} C_{11520}^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{11520-k} = \dots$$

Solution (Cont.)

考虑用近似计算, 利用 De Moivre-Laplace 中心极限定理,

$$\frac{n_A - 11520 \times 1/6}{\sqrt{11520 \times 1/6 \times 5/6}} \underset{\text{近似}}{\sim} N(0, 1).$$

所以

$$P\{n_A > 2160\} \approx 1 - \Phi\left(\frac{2160 - 11520 \times 1/6}{\sqrt{11520 \times 1/6 \times 5/6}}\right) = 1 - \Phi(6) = 0.$$

因此可以认为该骰子是不均匀的.

Example 13

设某工厂有 400 台同类机器, 各台机器发生故障的概率都是 0.02, 各台机器工作是相互独立的, 试求机器出故障的台数不小于 2 的概率.

Solution

设机器出故障的台数为 X , 则 $X \sim B(400, 0.02)$, 分别用三种方法计算:

① 用二项分布计算

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - 0.98^{400} - 400 \times 0.02 \times 0.98^{399} = 0.9972.$$

② 用泊松分布近似计算, $\lambda = np = 400 \times 0.02 = 8$,

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) \approx 1 - e^{-8} - 8e^{-8} = 0.9970.$$

Solution (Cont.)

3 用正态分布近似计算

$$\begin{aligned}\mathsf{P}(X \geq 2) &= 1 - \mathsf{P}(X \leq 1) \approx 1 - \Phi\left(\frac{1 - np}{\sqrt{np(1 - p)}}\right) \\ &= 1 - \Phi\left(\frac{1 - 400 \times 0.02}{\sqrt{400 \times 0.02 \times 0.98}}\right) \\ &= \Phi(2.5) = 0.9938.\end{aligned}$$

Example 14

设随机变量 X_1, X_2, \dots, X_{50} 相互独立同分布, $X_1 \sim U(-1, 1)$, 分别求 (1) $\frac{1}{50} \sum_{k=1}^{50} X_k$,
 (2) $\frac{1}{50} \sum_{k=1}^{50} |X_k|$, (3) $\frac{1}{50} \sum_{k=1}^{50} X_k^2$ 的近似分布.

Solution

由中心极限定理, $\frac{1}{50} \sum_{k=1}^{50} X_k$, $\frac{1}{50} \sum_{k=1}^{50} |X_k|$, $\frac{1}{50} \sum_{k=1}^{50} X_k^2$ 均近似服从正态分布.

因为, $E(X_1) = 0$, $\text{Var}(X_1) = \frac{1}{3}$, 所以 $\frac{1}{50} \sum_{k=1}^{50} X_k \stackrel{\text{近似}}{\sim} N\left(0, \frac{1}{150}\right)$.

Solution (Cont.)

因为

$$\mathbb{E}(|X_1|) = \frac{1}{2}, \quad \text{Var}(|X_1|) = \mathbb{E}(X_1^2) - [\mathbb{E}(|X_1|)]^2 = \frac{1}{12},$$

所以 $\frac{1}{50} \sum_{k=1}^{50} |X_k| \underset{\text{近似}}{\sim} N\left(\frac{1}{2}, \frac{1}{600}\right)$.

因为

$$\mathbb{E}(X_1^2) = \frac{1}{3}, \quad \text{Var}(X_1^2) = \mathbb{E}(X_1^4) - [\mathbb{E}(X_1^2)]^2 = \frac{1}{5} - \frac{1}{9} = \frac{4}{45},$$

所以 $\frac{1}{50} \sum_{k=1}^{50} X_k^2 \underset{\text{近似}}{\sim} N\left(\frac{1}{3}, \frac{2}{1125}\right)$.

Example 15

在 n 重 Bernoulli 试验中, 若已知每次试验事件 A 出现的概率为 0.75, 试利用中心极限定理,

- ① 若 $n = 7500$, 估计 A 出现的频率在 0.74 至 0.76 之间的概率至少有多大;
- ② 估计 n , 试 A 出现的频率在 0.74 至 0.76 之间的概率不小于 0.90.

Solution

设在 n 重 Bernoulli 试验中, 事件 A 出现的次数为 X_n , 则 $X_n \sim B(n, 0.75)$, 于是

$$\mathbb{E}(X_n) = np = 0.75n, \quad \text{Var}(X_n) = np(1 - p) = 0.1875n.$$

当 n 充分大时, $X_n \stackrel{\text{近似}}{\sim} N(0.75n, 0.1875n)$.

Solution (Cont.)

所以

$$\mathbb{P}\left\{0.74 < \frac{X_n}{n} < 0.76\right\} \approx \Phi\left(\frac{0.76n - 0.75n}{\sqrt{0.1875n}}\right) - \Phi\left(\frac{0.74n - 0.75n}{\sqrt{0.1875n}}\right) = 2\Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) - 1$$

① 已知 $n = 7500$, 因此此概率至少为 $2\Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) - 1 = 2\Phi(2) - 1 = 0.9544$.

② 若要求此概率不小于 0.90, 则 $2\Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) - 1 \geq 0.90$

$$\Rightarrow \Phi\left(\frac{0.04\sqrt{n}}{\sqrt{3}}\right) \geq 0.95 = \Phi(1.645) \Rightarrow n \geq (25 \times 1.645)^2 \times 3 = 5074.$$

(与 Chebyshev 不等式估计 $n \geq 18750$ 作比较.)

第五章的主要知识点:

- 依概率收敛的定义: 若对于 $\forall \varepsilon > 0$, 均有 $\lim_{n \rightarrow +\infty} P\{|Y_n - c| \geq \varepsilon\} = 0$ 成立, 则

$$Y_n \xrightarrow{P} c, \text{ 当 } n \rightarrow +\infty;$$

- 依概率收敛的性质;
- Markov 不等式和 Chebyshev 不等式;
- 大数律: 本书中的大数律均为:

在一定条件下, 有 $\frac{1}{n} \sum_{i=1}^n X_i - E(\frac{1}{n} \sum_{i=1}^n X_i) \xrightarrow{P} 0$, 当 $n \rightarrow +\infty$ 时.

需掌握: 贝努里大数律、辛钦大数律、Chebyshev 大数律的推论;

- 中心极限定理 (CLT): 独立同分布的 CLT、De Moivre-Laplace CLT.

谢谢!