第七讲大数定律

龚鹤扬

中国科学技术大学统计学博士
上海芯梯科技有限公司

2025 年 5 月 7 日

目录

1. 引言
2. 切比雪夫不等式
3. 弱大数定律 (WLLN)
4. 强大数定律 (SLLN)
5. 大数定律的应用
6. 总结与展望

引言：大数定律的重要性与局限

并非对所有随机变量都成立的普适规律

- 本讲主题：大数定律 (LLN) - 重要的极限定理。
- 核心思想：大量重复试验中，频率 $\rightarrow$ 概率；样本均值 $\rightarrow$ 总体期望。
- LLN：理论概率 $\leftrightarrow$ 统计推断的桥梁；多种统计方法的基础。
- 重要：LLN 结论依赖特定条件 (如期望存在、方差有限)。
- 条件不满足？LLN 可能失效。反例：柯西分布，探讨条件必要性。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个重要的概率上界

- 作用：给出随机变量偏离期望概率的上界。
- 特点：不依赖具体分布，仅需期望、方差存在且有限。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个重要的概率上界

- 作用：给出随机变量偏离期望概率的上界。
- 特点：不依赖具体分布，仅需期望、方差存在且有限。

定理 7.1 (切比雪夫不等式)

设随机变量 X 具有期望 $\mathrm{E}(X) = \mu$ 和方差 $\mathrm{Var}(X) = \sigma^2$ (其中 $0 < \sigma^2 < +\infty$)。则对于任意正数 $\epsilon > 0$，有：

$$\mathrm{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

或者等价地：

$$\mathrm{P}(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

- 作用：给出随机变量偏离期望概率的上界。
- 特点：不依赖具体分布，仅需期望、方差存在且有限。

定理 7.1 (切比雪夫不等式)

设随机变量 X 具有期望 $\mathrm{E}(X) = \mu$ 和方差 $\mathrm{Var}(X) = \sigma^2$ (其中 $0 < \sigma^2 < +\infty$)。则对于任意正数 $\epsilon > 0$，有：

$$\mathrm{P}(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$

或者等价地：

$$\mathrm{P}(|X - \mu| < \epsilon) \geq 1 - \frac{\sigma^2}{\epsilon^2}$$

意义与特点：

- 普适性：条件满足时，对任何分布成立。
- 实用性：分布未知时的粗略估计。
- 局限性：界限通常较宽松。

7.2 弱大数定律 (Weak Law of Large Numbers, WLLN)

样本均值的依概率收敛

WLLN 核心：特定条件下，样本均值 依概率收敛 于总体期望。

定理 7.2 (切比雪夫弱大数定律)

设 $X_1, X_2, \ldots, X_n, \ldots$ 是一列相互独立、具有相同期望 $\mathrm{E}(X_i) = \mu$ 和相同有限方差 $\mathrm{Var}(X_i) = \sigma^2 < +\infty$ 的随机变量序列。令样本均值为 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。则对于任意 $\epsilon > 0$，有：

$$\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

或者等价地：

$$\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| < \epsilon) = 1$$

这表示样本均值 $\bar{X}_n$ 依概率收敛于 μ，记作 $\bar{X}_n \xrightarrow{P} \mu$。

注：依赖方差有限；由切比雪夫不等式证明。

定理 7.3 (伯努利弱大数定律)

设 n_A 是 n 次独立重复伯努利试验中事件 A 发生的次数，p 是事件 A 在每次试验中发生的概率。令 $f_n = \frac{n_A}{n}$ 为事件 A 发生的频率。则对于任意 $\epsilon > 0$，有：

$$\lim_{n\to\infty} \mathrm{P}\left(|f_n - p| \geq \epsilon\right) = 0$$

揭示了"频率稳定性" 本质：n 很大时，f_n 高概率接近 p。

注：切比雪夫 $WLLN$ 特例；频率估计概率的理论依据。

期望存在的重要性：引入柯西分布

一个大数定律不成立的著名反例

柯西分布 (Cauchy Distribution)

PDF:

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x-x_0}{\gamma}\right)^2\right]}$$

x_0: 位置参数 (峰值、中位数、众数); $\gamma > 0$: 尺度参数 (宽度、半峰全宽)。

标准柯西 ($x_0 = 0, \gamma = 1$): $f(x) = \frac{1}{\pi(1+x^2)}$。

特点:

- 钟形对称，但重尾 (比正态分布尾部厚)。
- 期望 $\mathrm{E}(X)$ 不存在 (积分发散)。
- 方差 $\mathrm{Var}(X)$ 亦不存在。

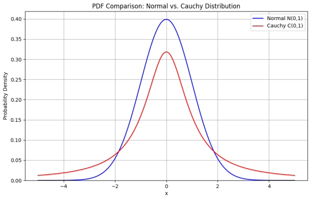

图: *

PDFs: N(0,1) vs C(0,1)

对 LLN 的影响:

- 不满足 LLN 的" 期望存在/方差有限" 条件。
- i.i.d. 柯西序列的样本均值不收敛。

可视化：正态 vs. 柯西样本均值路径

样本均值随样本量 n 变化的轨迹

正态分布 $N(0,1)$
$(\mathrm{E}(X)=0,\ \mathrm{Var}(X)=1)$

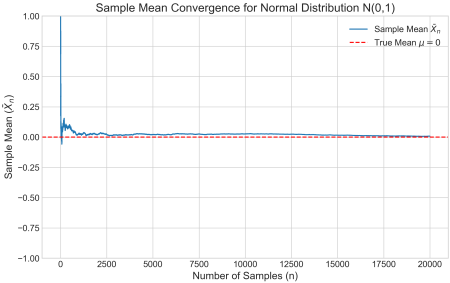

图: $\bar{X}_n$ 快速稳定收敛到 $\mu = 0$

柯西分布 $C(0,1)$
$(\mathrm{E}(X)$ 不存在$)$

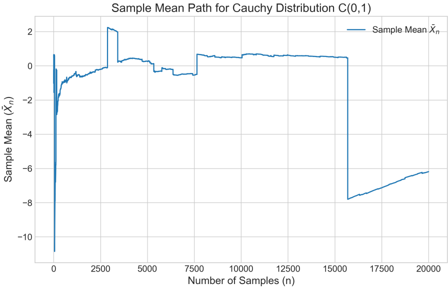

图: $\bar{X}_n$ 持续剧烈波动, 不收敛

柯西样本均值：持续波动，不收敛 (极端值影响)

辛钦弱大数定律 (Khinchin's WLLN)

期望存在即可保证依概率收敛 (i.i.d. 情况)

辛钦 WLLN：i.i.d. 序列，仅要求期望存在 (无需方差有限)。

定理 7.4 (辛钦弱大数定律)

设 $X_1, X_2, \ldots, X_n, \ldots$ 是一列独立同分布 (i.i.d.) 的随机变量序列，且其共同的期望 $\mathrm{E}(X_i) = \mu$ 存在 (即 $|\mu| < \infty$)。令样本均值为 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。则对于任意 $\epsilon > 0$，有：

$$\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$$

即 $\bar{X}_n \xrightarrow{P} \mu$。

柯西分布与辛钦 WLLN：

- i.i.d. 柯西序列：期望不存在 $\implies$ 不满足辛钦 WLLN 条件。
- 结论：其样本均值 $\bar{X}_n$ 不依概率收敛。

7.3 强大数定律 (Strong Law of Large Numbers, SLLN)

样本均值的几乎必然收敛 (更强的收敛)

SLLN：更强收敛性，样本均值几乎必然收敛 ($\xrightarrow{a.s.}$) 于总体期望 (同样要求期望存在)。

定理 7.5 (柯尔莫哥洛夫强大数定律)

设 $X_1, X_2, \ldots, X_n, \ldots$ 是一列独立同分布 (i.i.d.) 的随机变量序列，且其共同的期望 $\mathrm{E}(X_i) = \mu$ 存在 (即 $|\mu| < \infty$) 。令样本均值为 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。则：

$$\mathrm{P}\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

这表示样本均值 $\bar{X}_n$ 几乎处处收敛 (Almost Surely Converge) 于 μ，记作 $\bar{X}_n \xrightarrow{a.s.} \mu$。

柯西分布与 SLLN：

- i.i.d. 柯西序列：期望不存在 $\Longrightarrow$ 不满足SLLN 条件。
- 结论：其样本均值 $\bar{X}_n$ 不几乎必然收敛。

注：柯尔莫哥洛夫强大数定律的条件与辛钦弱大数定律相同 (*i.i.d.* 且期望存在)，但结论更强。

核心区别：

- **WLLN - 依概率收敛 ($\xrightarrow{P}$)**: $\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$. 关注概率的极限。
- **SLLN - 几乎必然收敛 ($\xrightarrow{a.s.}$)**: $\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. 关注样本路径的极限。

WLLN 与 SLLN 的区别与联系 (1)：正态分布

依概率收敛 vs. 几乎必然收敛

核心区别：

- **WLLN** - **依概率收敛** **($\xrightarrow{P}$)**: $\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$. 关注概率的极限。
- **SLLN** - **几乎必然收敛** **($\xrightarrow{a.s.}$)**: $\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$. 关注样本路径的极限。

条件与对比：

- SLLN 比 WLLN 更强 ($\xrightarrow{a.s.} \Longrightarrow \xrightarrow{P}$)。
- 两者最广泛应用的 i.i.d. 版本都要求 期望存在 ($|\mu| < \infty$)。
- **对于正态分布 (例如 $N(\mu, \sigma^2)$):**
  - 期望 μ 存在、方差 σ^2 有限 $\Longrightarrow$ 满足 LLN 条件。
  - i.i.d. 正态样本均值 $\bar{X}_n$: $\xrightarrow{P} \mu$ 且 $\xrightarrow{a.s.} \mu$。

WLLN 与 SLLN 的区别与联系 (2)：柯西分布反例

期望不存在导致大数定律失效

特性	正态分布	柯西分布
均值	存在 (μ)	不存在 (位置参数 x_0)
方差	有限 (σ^2)	不存在 (尺度参数 γ)
LLN	成立	不成立
样本均值行为	收敛至 μ	剧烈波动

表：正态分布与柯西分布对比

WLLN 与 SLLN 的区别与联系 (2)：柯西分布反例

期望不存在导致大数定律失效

特性	正态分布	柯西分布
均值	存在 (μ)	不存在 (位置参数 x_0)
方差	有限 (σ^2)	不存在 (尺度参数 γ)
LLN	成立	不成立
样本均值行为	收敛至 μ	剧烈波动

表：正态分布与柯西分布对比

关键特性：对于 i.i.d. 柯西序列的样本均值 $\bar{X}_n$

- 期望不存在 $\implies$ 不满足 LLN 的期望存在条件，即

$$\bar{X}_n \stackrel{P}{\not\to} \text{const}, \quad \bar{X}_n \stackrel{a.s.}{\not\longrightarrow} \text{const}$$

- 惊人事实：若 $X_i \sim C(0,1)$ i.i.d.，则 $\bar{X}_n = \frac{1}{n}\sum X_i \sim C(0,1)$ (仍为标准柯西分布，与 n 无关!)

7.4 大数定律的应用 (Applications of LLN)

应用广泛，但基础是满足 LLN 条件 (尤其期望存在)

- **前提**：应用均假设满足 LLN 条件 (尤其期望存在)。
- **理论与解释**：
  - 频率稳定性 (伯努利 LLN)。
  - 解释平均结果的稳定趋向。
- **统计推断**：
  - 参数估计 (样本均值 $\to$ 总体期望，前提：期望存在!)。
  - 蒙特卡洛方法 (随机抽样估计)。
- **风险管理与保险**：预测平均赔付。
- **质量控制**：样本推断总体。
- **物理与工程**：多次测量取平均减小误差。
  - 警示：若误差为柯西分布，平均无效!

核心警示：LLN 条件 (尤其期望存在) 是应用有效的关键。

总结与展望

本讲小结

- **切比雪夫不等式**: 概率上界 (需方差)。
- **弱大数定律 (WLLN)**: $\bar{X}_n \xrightarrow{P} \mu$ (如辛钦: i.i.d., $\mathrm{E}\,X_i$ 存在)。
- **强大数定律 (SLLN)**: $\bar{X}_n \xrightarrow{a.s.} \mu$ (如柯尔莫哥洛夫: i.i.d., $\mathrm{E}\,X_i$ 存在); SLLN $\implies$ WLLN.
- LLN 条件核心: **期望存在性。**
- 反例 - 柯西分布: $\mathrm{E}\,X$ 不存在 $\implies$ LLN 不适用; $\bar{X}_n$ 仍为柯西。
- **LLN 地位**: 连接概率论与统计实践的核心桥梁。

展望

- 大数定律描述了平均行为的极限。
- 核心极限定理 中心极限定理 (CLT)：描述样本均值的分布 $\to$ 正态分布 (通常需 $\mathrm{E}\,X, \mathrm{Var}\,X$ 存在)。
- 大数定律和中心极限定理共同构成了现代统计推断的理论基石。

谢谢聆听！

相关数学证明参考资料请访问:
https://1587causalai.github.io/BasicProbabilityLectureNotes

问题与讨论