第七讲大数定律

正态分布的普遍性与柯西分布的特殊性

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. 大数定律的应用与警示
6. 总结与展望

引言：大数定律的重要性与适用性

随机现象中的规律性

- 本讲讨论概率论中一组重要的极限定理——大数定律 (Law of Large Numbers, LLN)。
- 核心思想：在大量重复试验中，事件发生的频率近似于其概率，样本均值收敛于总体期望。
- 大数定律是连接理论概率与统计推断的桥梁。
- 重要前提：大数定律的结论依赖于随机变量满足特定的条件（如期望存在、方差有限等）。
- 我们将以常见的正态分布为例，说明大数定律如何在其条件下成立；并以柯西分布为例，探讨当条件不满足时，大数定律为何失效。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个普适的概率上界

- 切比雪夫不等式给出了随机变量偏离其期望值的概率的一个上界。
- 这个界不依赖于随机变量的具体分布形式，只需要其期望和方差存在且有限。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个普适的概率上界

- 切比雪夫不等式给出了随机变量偏离其期望值的概率的一个上界。
- 这个界不依赖于随机变量的具体分布形式，只需要其期望和方差存在且有限。

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：

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

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个普适的概率上界

- 切比雪夫不等式给出了随机变量偏离其期望值的概率的一个上界。
- 这个界不依赖于随机变量的具体分布形式，只需要其期望和方差存在且有限。

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：

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

示例：正态分布

- 对于正态分布 $N(\mu, \sigma^2)$，其期望为 μ，方差为 σ^2 (均存在且有限)。
- 因此，切比雪夫不等式适用于正态分布，为 $|X - \mu|$ 的偏离提供了一个（尽管可能宽松的）概率上界。

7.2 弱大数定律 (WLLN): 切比雪夫方法

基于方差存在的依概率收敛

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:

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

7.2 弱大数定律 (WLLN): 切比雪夫方法

基于方差存在的依概率收敛

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：

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

示例：正态分布序列

- 若 $X_i \sim N(\mu, \sigma^2)$ 且相互独立，则它们满足切比雪夫 WLLN 的所有条件 (独立、同期望 μ、同有限方差 σ^2)。
- 因此，i.i.d. 正态随机变量的样本均值 $\bar{X}_n$ 依概率收敛于总体期望 μ。

注：此定理依赖于方差存在且有限，并通过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 $\epsilon > 0$: $\lim_{n\to\infty} \mathrm{P}\left(|f_n - p| \geq \epsilon\right) = 0$。

注：这是切比雪夫 *WLLN* 的特例，因为伯努利变量 $X_i \sim B(1, p)$ 的 $\mathrm{E}(X_i) = p$ 且 $\mathrm{Var}(X_i) = p(1-p)$ 有限。

大数定律的边界：引入柯西分布

当期望不再可靠

- 切比雪夫 WLLN 要求方差有限。许多其他版本的大数定律（如辛钦 WLLN，柯尔莫哥洛夫 SLLN）放宽了对方差的要求，但一个更核心的条件是：期望的存在性。

大数定律的边界：引入柯西分布

当期望不再可靠

- 切比雪夫 WLLN 要求方差有限。许多其他版本的大数定律（如辛钦 WLLN，柯尔莫哥洛夫 SLLN）放宽了对方差的要求，但一个更核心的条件是：期望的存在性。
- **思考**：如果一个分布连期望都没有，样本均值会收敛到哪里呢?
- 柯西分布 (Cauchy Distribution) 就是这样一个期望不存在的著名例子。

大数定律的边界：引入柯西分布

当期望不再可靠

- 切比雪夫 WLLN 要求方差有限。许多其他版本的大数定律（如辛钦 WLLN，柯尔莫哥洛夫 SLLN）放宽了对方差的要求，但一个更核心的条件是：期望的存在性。
- **思考**：如果一个分布连期望都没有，样本均值会收敛到哪里呢？
- 柯西分布 (Cauchy Distribution) 就是这样一个期望不存在的著名例子。

柯西分布 $C(x_0, \gamma)$

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

- x_0：位置参数 (中位数)，γ：尺度参数。
- 标准柯西 $C(0,1)$ $(x_0 = 0, \gamma = 1)$：$f(x) = \frac{1}{\pi(1+x^2)}$。
- 特点：重尾，期望 $\mathrm{E}(X)$ 不存在，方差 $\mathrm{Var}(X)$ 无限大。

大数定律的边界：引入柯西分布

当期望不再可靠

- 切比雪夫 WLLN 要求方差有限。许多其他版本的大数定律（如辛钦 WLLN，柯尔莫哥洛夫 SLLN）放宽了对方差的要求，但一个更核心的条件是：期望的存在性。
- **思考**：如果一个分布连期望都没有，样本均值会收敛到哪里呢？
- 柯西分布 (Cauchy Distribution) 就是这样一个期望不存在的著名例子。

柯西分布 $C(x_0, \gamma)$

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

- x_0：位置参数 (中位数), γ：尺度参数。
- 标准柯西 $C(0,1)$ $(x_0 = 0, \gamma = 1)$：$f(x) = \frac{1}{\pi(1+x^2)}$。
- 特点：重尾，期望 $\mathrm{E}(X)$ 不存在，方差 $\mathrm{Var}(X)$ 无限大。

这意味着柯西分布不满足切比雪夫 *WLLN* 的方差条件，也不满足后续 *WLLN*/*SLLN* 的期望存在条件。

可视化：正态 vs. 柯西样本均值路径

大数定律的生效与失效

观察 i.i.d. 序列样本均值 $\bar{X}_n$ 随样本量 n 增大的典型路径：

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

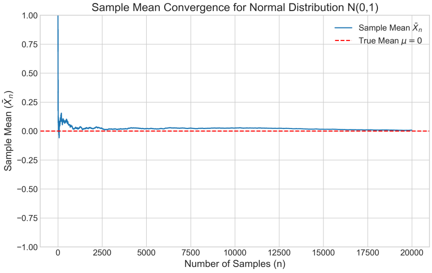

图: $\bar{X}_n$ 快速稳定收敛到 $\mu = 0$

柯西分布 $C(0,1)$
$(\mathrm{E}(X)$ 不存在$)$

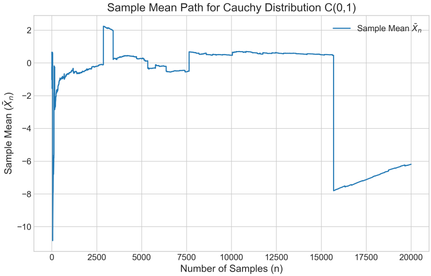

图: $\bar{X}_n$ 持续剧烈波动, 不收敛

启示:

- 正态分布的样本均值行为良好，符合大数定律预期。
- 柯西分布的极端值持续" 污染" 样本均值，使其无法稳定。

WLLN (续): 辛钦弱大数定律

期望存在即可 (i.i.d. 情况)

定理 7.4 (辛钦弱大数定律)

设 $X_1, X_2, \ldots$ 是一列独立同分布 (i.i.d.) 随机变量，其共同期望 $\mathrm{E}(X_i) = \mu$ 存在 ($|\mu| < \infty$)。则样本均值 $\bar{X}_n = \frac{1}{n}\sum X_i \xrightarrow{P} \mu$。

WLLN (续): 辛钦弱大数定律
期望存在即可 (i.i.d. 情况)

定理 7.4 (辛钦弱大数定律)

设 $X_1, X_2, \ldots$ 是一列独立同分布 (i.i.d.) 随机变量，其共同期望 $\mathrm{E}(X_i) = \mu$ 存在 $(|\mu| < \infty)$。则样本均值 $\bar{X}_n = \frac{1}{n}\sum X_i \xrightarrow{P} \mu$。

对比应用：

- 正态分布 $N(\mu, \sigma^2)$：
  - i.i.d.，期望 μ 存在。
  - $\Longrightarrow$ 满足辛钦 WLLN 条件，$\bar{X}_n \xrightarrow{P} \mu$。(无需有限方差)
- 柯西分布 $C(x_0, \gamma)$：
  - i.i.d.，但期望不存在。
  - $\Longrightarrow$ 不满足辛钦 WLLN 条件，其结论不适用。

7.3 强大数定律 (SLLN): 柯尔莫哥洛夫

更强的几乎必然收敛 (i.i.d. 情况)

定理 7.5 (柯尔莫哥洛夫强大数定律)

设 $X_1, X_2, \dots$ 是一列独立同分布 (i.i.d.) 随机变量，其共同期望 $\mathrm{E}(X_i) = \mu$ 存在 $(|\mu| < \infty)$。则样本均值 $\bar{X}_n = \frac{1}{n}\sum X_i \xrightarrow{a.s.} \mu$。$(\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1)$

7.3 强大数定律 (SLLN): 柯尔莫哥洛夫

更强的几乎必然收敛 (i.i.d. 情况)

定理 7.5 (柯尔莫哥洛夫强大数定律)

设 $X_1, X_2, \ldots$ 是一列独立同分布 (i.i.d.) 随机变量，其共同期望 $\mathrm{E}(X_i) = \mu$ 存在 $(|\mu| < \infty)$。则样本均值 $\bar{X}_n = \frac{1}{n}\sum X_i \xrightarrow{a.s.} \mu$。($\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$)

对比应用：

- **正态分布** $N(\mu, \sigma^2)$：
  - i.i.d.，期望 μ 存在。
  - $\Longrightarrow$ 满足柯尔莫哥洛夫 SLLN 条件，$\bar{X}_n \xrightarrow{a.s.} \mu$。(结论更强)
- **柯西分布** $C(x_0, \gamma)$：
  - i.i.d.，但期望不存在。
  - $\Longrightarrow$ 不满足SLLN 条件，其结论不适用。

SLLN 条件与辛钦 WLLN 相同 (i.i.d. 且期望存在)，但结论更强 (几乎必然收敛)。

WLLN 与 SLLN 的区别与联系

依概率收敛 ($\xrightarrow{P}$) vs. 几乎必然收敛 ($\xrightarrow{a.s.}$)

WLLN ($\xrightarrow{P}$):

- $\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$
- 关注概率：当 n 大时，$\bar{X}_n$ 偏离 μ 很远的可能性很小。
- 不排除在某些 n 处，$\bar{X}_n$ 可能偏离 μ 较远，但这种事件发生的概率随 n 趋于 0。

SLLN ($\xrightarrow{a.s.}$):

- $\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$
- 关注样本路径：对于几乎所有的随机试验序列，$\bar{X}_n$ 的极限就是μ。
- 意味着除了一个概率为 0 的例外集合，$\bar{X}_n$ 最终会" 停在"μ 处。

WLLN 与 SLLN 的区别与联系

依概率收敛 ($\xrightarrow{P}$) vs. 几乎必然收敛 ($\xrightarrow{a.s.}$)

WLLN ($\xrightarrow{P}$):

- $\lim_{n\to\infty} \mathrm{P}(|\bar{X}_n - \mu| \geq \epsilon) = 0$
- 关注概率：当 n 大时，$\bar{X}_n$ 偏离 μ 很远的可能性很小。
- 不排除在某些 n 处，$\bar{X}_n$ 可能偏离 μ 较远，但这种事件发生的概率随 n 趋于 0。

SLLN ($\xrightarrow{a.s.}$):

- $\mathrm{P}(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$
- 关注样本路径：对于几乎所有的随机试验序列，$\bar{X}_n$ 的极限就是μ。
- 意味着除了一个概率为 0 的例外集合，$\bar{X}_n$ 最终会" 停在"μ 处。

强度与适用： SLLN $\implies$ WLLN (几乎必然收敛更强)。

- **正态分布** $N(\mu, \sigma^2)$ **(i.i.d.):** 期望 μ 存在。
  - $\bar{X}_n \xrightarrow{P} \mu$ (由辛钦 WLLN)
  - $\bar{X}_n \xrightarrow{a.s.} \mu$ (由柯尔莫哥洛夫 SLLN)
- **柯西分布** $C(x_0, \gamma)$ **(i.i.d.):** 期望不存在。
  - $\bar{X}_n$ 不依概率收敛，也不几乎必然收敛到任何常数。
  - 惊人事实：若 $X_i \sim C(0,1)$ i.i.d.，则 $\bar{X}_n \sim C(0,1)$ 对所有 n 都成立!

大数定律是概率论与数理统计的重要基石：

- **频率稳定性**：为用频率近似概率提供了理论依据 (伯努利 WLLN)。
- **参数估计**：样本均值 $\bar{X}_n$ 是总体期望 μ 的一致估计量。
  - 例如：对于正态总体 $N(\mu, \sigma^2)$，用 $\bar{X}_n$ 估计 μ 是合理的。
  - 警示：若总体为柯西分布，$\bar{X}_n$ 不是其位置参数 x_0 的一致估计量 *(*因 μ 不存在，*LLN* 不适用*)*。用中位数估计 x_0 更稳健。
- **蒙特卡洛方法**：通过大量随机抽样，用样本均值估计复杂积分或期望。
  - 适用于被积函数对应随机变量期望存在的情况。
- **风险管理与保险**：预测平均赔付，依赖于大量事件下平均损失的稳定性。
- **物理与工程**：多次测量的平均值通常比单次测量更接近真实值。
  - 警示：如果测量误差服从柯西分布，增加测量次数并取平均无益于提高精度！

核心：大数定律的应用有效性，强依赖于其前提条件（尤其是期望存在）是否满足。

总结与展望

本讲小结：大数定律的精髓与边界

- **切比雪夫不等式**：提供了概率的通用上界（需期望、方差存在）。
- **大数定律**：描述样本均值在特定条件下收敛于总体期望。
  - WLLN (依概率 $\xrightarrow{P}$)，SLLN (几乎必然 $\xrightarrow{a.s.}$)。SLLN 更强。
  - 关键版本 (辛钦 WLLN, 柯尔莫哥洛夫 SLLN) 要求 i.i.d. 且期望存在。
- **正态分布 (i.i.d., $\mathrm{E}(X) = \mu$ 存在)**:
  - 样本均值 $\bar{X}_n \xrightarrow{P} \mu$ 且 $\bar{X}_n \xrightarrow{a.s.} \mu$。大数定律完美适用。
- **柯西分布 (i.i.d., $\mathrm{E}(X)$ 不存在)**:
  - 样本均值 $\bar{X}_n$ 不收敛。大数定律不适用。
  - $\bar{X}_n$ 仍服从柯西分布，揭示了期望不存在时的特殊性。
- 理解大数定律的条件和局限性对正确应用至关重要。

总结与展望

本讲小结：大数定律的精髓与边界

- **切比雪夫不等式**：提供了概率的通用上界（需期望、方差存在）。
- **大数定律**：描述样本均值在特定条件下收敛于总体期望。
  - WLLN (依概率 $\xrightarrow{P}$)，SLLN (几乎必然 $\xrightarrow{a.s.}$)。SLLN 更强。
  - 关键版本 (辛钦 WLLN, 柯尔莫哥洛夫 SLLN) 要求 i.i.d. 且期望存在。
- **正态分布 (i.i.d., $\mathrm{E}(X) = \mu$ 存在)**:
  - 样本均值 $\bar{X}_n \xrightarrow{P} \mu$ 且 $\bar{X}_n \xrightarrow{a.s.} \mu$。大数定律完美适用。
- **柯西分布 (i.i.d., $\mathrm{E}(X)$ 不存在)**:
  - 样本均值 $\bar{X}_n$ 不收敛。大数定律不适用。
  - $\bar{X}_n$ 仍服从柯西分布，揭示了期望不存在时的特殊性。
- 理解大数定律的条件和局限性对正确应用至关重要。

展望：中心极限定理 (CLT)

- 大数定律告诉我们样本均值收敛到哪里。
- 下一步，中心极限定理 (CLT) 将揭示：当 n 很大时，样本均值围绕总体期望的波动形态近似于何种分布 (通常是正态分布)。
- CLT 也有其前提条件，如经典的 Lindeberg-Lévy CLT 要求 i.i.d. 且具有有限的期望和方差。这对正态分布适用，但对柯西分布 (因方差无限) 则不适用标准 CLT。

谢谢聆听！ 问题与讨论