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 6 日

目录

引言：大数定律的重要性

- 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。
- 它为许多统计方法的合理性提供了理论基础。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个重要的概率上界

- 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)

一个重要的概率上界

- 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 E(X) $= \mu$ 和方差 D(X) $= \sigma^2$ (其中 $0 < \sigma^2 < +\infty$)。则对于任意正数 $\epsilon > 0$，有:

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

或者等价地:

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

7.1 切比雪夫不等式 (Chebyshev's Inequality)

一个重要的概率上界

- 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 E(X) $= \mu$ 和方差 D(X) $= \sigma^2$ (其中 $0 < \sigma^2 < +\infty$)。则对于任意正数 $\epsilon > 0$，有:

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

或者等价地:

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

意义与特点:

- 普适性: 对任何分布都成立（只要期望、方差存在）。
- 实用性: 在分布未知时，可以提供一个粗略的估计。
- 局限性: 通常这个界比较宽松，即估计可能不够精确。

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

样本均值的依概率收敛

弱大数定律 (WLLN) 表明，当试验次数 n 足够大时，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 $\mathrm{D}(X_i) = \sigma^2 < +\infty$ 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：

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

或者等价地：

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

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

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：

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

或者等价地：

$$\lim_{n\to\infty} P(|f_n - p| < \epsilon) = 1$$

伯努利大数定律是切比雪夫弱大数定律的特例，它揭示了“频率稳定性”的本质：当试验次数 n 很大时，事件的频率 f_n 会以很高的概率接近其真实的概率 p。

注：伯努利大数定律是最早提出的大数定律形式，为用频率估计概率提供了理论依据。

辛钦弱大数定律 (Khinchin's WLLN)

更宽松的方差条件

辛钦弱大数定律对随机变量序列的方差要求更低。

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 $\mathsf{E}(X_i) = \mu$ 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：

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

即 $\bar{X}_n \xrightarrow{P} \mu$。

注意：辛钦 WLLN 不要求方差存在或有限，仅要求期望存在且序列独立同分布。这使得它比切比雪夫 WLLN 的适用范围更广。

强大数定律 (SLLN) 比弱大数定律具有更强的收敛性，它表明样本均值几乎必然收敛 (或称以概率 1 收敛) 于总体期望。

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:

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

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

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，但结论更强。

WLLN 与 SLLN 的区别与联系

依概率收敛 vs. 几乎必然收敛

核心区别:

- **弱大数定律 (WLLN) - 依概率收敛 ($\xrightarrow{P}$)**:
  - $\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$
  - 含义：对于任意小的 $\epsilon > 0$ 和 $\delta > 0$，存在一个 N，当 $n > N$ 时，$P(|\bar{X}_n - \mu| < \epsilon) > 1 - \delta$。
  - 描述的是当 n 足够大时，$\bar{X}_n$ 单次观察值落在 μ 的小邻域内的概率趋近于 1。可能存在某些 n 值，$\bar{X}_n$ 偏离 μ 较远，但这种偏离发生的概率随着 n 增大而减小。
- **强大数定律 (SLLN) - 几乎必然收敛 ($\xrightarrow{a.s.}$)**:
  - $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$
  - 含义：对于几乎所有的样本序列 (即除去一个概率为 0 的例外集合)，当 $n \to \infty$ 时，$\bar{X}_n$ 的极限值就是 μ。
  - 描述的是整个序列 $\bar{X}_1, \bar{X}_2, \ldots$ 的收敛行为。

WLLN 与 SLLN 的区别与联系

依概率收敛 vs. 几乎必然收敛

核心区别:

- **弱大数定律 (WLLN) - 依概率收敛 ($\xrightarrow{P}$)**:
  - $\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$
  - **含义**：对于任意小的 $\epsilon > 0$ 和 $\delta > 0$，存在一个 N，当 $n > N$ 时，$P(|\bar{X}_n - \mu| < \epsilon) > 1 - \delta$。
  - 描述的是当 n 足够大时，$\bar{X}_n$ 单次观察值落在 μ 的小邻域内的概率趋近于 1。可能存在某些 n 值，$\bar{X}_n$ 偏离 μ 较远，但这种偏离发生的概率随着 n 增大而减小。
- **强大数定律 (SLLN) - 几乎必然收敛 ($\xrightarrow{a.s.}$)**:
  - $P(\lim_{n\to\infty} \bar{X}_n = \mu) = 1$
  - **含义**：对于几乎所有的样本序列 (即除去一个概率为 0 的例外集合)，当 $n \to \infty$ 时，$\bar{X}_n$ 的极限值就是 μ。
  - 描述的是整个序列 $\bar{X}_1, \bar{X}_2, \ldots$ 的收敛行为。

联系与强度:

- SLLN $\implies$ WLLN：几乎必然收敛是比依概率收敛更强的收敛模式。
- **条件**：SLLN 通常需要与 WLLN 相似或略强的条件 (如柯尔莫哥洛夫 SLLN 和辛钦 WLLN 的条件都是 i.i.d. 和期望存在)。切比雪夫 WLLN 条件不同 (独立、同期望、同有限方差)。

7.4 大数定律的应用 (Applications of LLN)

大数定律是概率论与数理统计之间重要的理论基石，应用广泛：

- **理论基础与解释现象**：
  - 频率的稳定性：为用频率近似概率提供了坚实的理论依据 (伯努利大数定律)。
  - 解释了在大量重复试验中，随机事件的平均结果会趋于一个稳定的值。
- **统计推断**：
  - 参数估计：样本均值是总体期望的一致估计量。这是矩估计法等参数估计方法的基础。
  - 蒙特卡洛方法：通过大量随机抽样来模拟复杂过程，估计积分、期望等数值解。例如，用投点法计算 π 的值。
- **风险管理与保险**：
  - 保险公司利用大数定律来预测在大量保单中可能发生的赔付总额，从而制定合理的保费，确保经营稳定。
- **质量控制与抽样检验**：
  - 通过对产品进行抽样检查，根据样本的合格率来推断整批产品的质量水平。
- **物理与工程**：
  - 测量误差的分析：多次测量的平均值通常比单次测量更接近真实值。

总结与展望

本讲小结

- **切比雪夫不等式**：提供了概率的通用上界。
- **弱大数定律 (WLLN)**：样本均值依概率收敛于总体期望。
    - 切比雪夫 WLLN (独立, 同期望, 同有限方差)
    - 伯努利 WLLN (频率稳定性)
    - 辛钦 WLLN (i.i.d., 期望存在)
- **强大数定律 (SLLN)**：样本均值几乎必然收敛于总体期望 (i.i.d., 期望存在)。
- SLLN 比 WLLN 更强，提供了关于整个样本路径收敛性的保证。
- 大数定律是连接概率论与统计实践的核心桥梁。

总结与展望

本讲小结

- **切比雪夫不等式**：提供了概率的通用上界。
- **弱大数定律 (WLLN)**：样本均值依概率收敛于总体期望。
  - 切比雪夫 WLLN (独立, 同期望, 同有限方差)
  - 伯努利 WLLN (频率稳定性)
  - 辛钦 WLLN (i.i.d., 期望存在)
- **强大数定律 (SLLN)**：样本均值几乎必然收敛于总体期望 (i.i.d., 期望存在)。
- SLLN 比 WLLN 更强，提供了关于整个样本路径收敛性的保证。
- 大数定律是连接概率论与统计实践的核心桥梁。

展望

- 大数定律有多种形式，对应不同的条件和收敛强度。
- 除了大数定律，概率论中的另一个核心极限定理是中心极限定理 (Central Limit Theorem, CLT)，它描述了样本均值的分布形态，将在后续课程中讨论。
- 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！ 问题与讨论