第七讲大数定律

概率论的基石：从理论到应用

龚鹤扬

上海芯梯科技有限公司

2025 年 5 月 6 日

本讲纲要

引言：大数定律的直观与重要性

- 本章讨论概率论中一组核心的极限定理——大数定律。

引言：大数定律的直观与重要性

- 本章讨论概率论中一组核心的极限定理——大数定律。
- 描述了在大量重复独立试验中，事件发生的频率如何逼近其概率，或者样本均值如何收敛于总体期望的现象。

引言：大数定律的直观与重要性

- 本章讨论概率论中一组核心的极限定理——大数定律。
- 描述了在大量重复独立试验中，事件发生的频率如何逼近其概率，或者样本均值如何收敛于总体期望的现象。
- 是连接 理论概率 与 统计推断 的关键桥梁。

引言：大数定律的直观与重要性

- 本章讨论概率论中一组核心的极限定理——大数定律。
- 描述了在大量重复独立试验中，事件发生的频率如何逼近其概率，或者样本均值如何收敛于总体期望的现象。
- 是连接 理论概率 与 统计推断 的关键桥梁。
- 为许多统计方法的合理性提供了坚实的理论基础。

7.1 切比雪夫不等式 (Chebyshev's Inequality)

- 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)

- 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。

切比雪夫不等式

设随机变量 X 具有期望 E(X) = 和方差 D(X) = ^{2} (其中 0 < ^{2} < +∞)。则对于任意正数 > 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)

- 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。

切比雪夫不等式

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

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

或者等价地：

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

核心特点：

- 普适性强：对任何分布都成立（只要期望方差存在）。
- 界限较松：实际应用中，这个上界往往比较宽松。

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

弱大数定律表明，当试验次数足够大时，样本均值依概率收敛于总体期望。

定理 (切比雪夫弱大数定律)

设 X₁, X₂, ..., X_n, ... 是一列相互独立、具有相同期望 E(X_i) = 和相同有限方差 D(X_i) = ² < +∞ 的随机变量序列。令 S_n = Σ_{i=1}ⁿ X_i 为前 n 个随机变量之和，$\bar{X}_n = S_n/n$ 为样本均值。则对于任意 > 0，有：

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

或者等价地：

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

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

伯努利弱大数定律

这是切比雪夫弱大数定律在伯努利试验下的一个重要特例。

定理 (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。则对于任意 > 0，有：

$$\lim_{n \to \infty} P\left(\left| \frac{n_A}{n} - p \right| < \epsilon \right) = 1$$

此定律说明了频率的稳定性：当试验次数 n 很大时，事件的频率 n_A/n 会以很高的概率接近其真实的概率 p。

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

强大数定律比弱大数定律描述了更强的收敛性，它表明样本均值几乎必然收敛于总体期望。

定理 (柯尔莫哥洛夫强大数定律)

设 X₁, X₂, ..., X_n, ... 是一列独立同分布 (i.i.d.) 的随机变量序列，且其共同期望 E(X_i) = 存在 (不要求方差存在或有限)。令 $\bar{X}_n = (X_1 + ... + X_n)/n$。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) 或以概率 1 收敛于 ，记作 $\bar{X}_n \xrightarrow{a.s.} \mu$。

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

强大数定律比弱大数定律描述了更强的收敛性，它表明样本均值几乎必然收敛于总体期望。

定理 (柯尔莫哥洛夫强大数定律)

设 X₁, X₂, ..., X_n, ... 是一列独立同分布 (i.i.d.) 的随机变量序列，且其共同期望 E(X_i) = 存在 (不要求方差存在或有限)。令 $\bar{X}_n = (X_1 + ... + X_n)/n$。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) 或以概率 1 收敛于 ，记作 $\bar{X}_n \xrightarrow{a.s.} \mu$。

注意

SLLN 的条件 (i.i.d. 且期望存在) 与切比雪夫 WLLN 的条件 (独立、同

WLLN 与 SLLN 的对比

核心区别：

- **收敛模式：**
  - WLLN (依概率收敛 $\xrightarrow{P}$)：

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

指对于任意小的 ϵ 和 δ，总存在 N，当 $n > N$ 时，$\bar{X}_n$ 落在 $(\mu - \epsilon, \mu + \epsilon)$ 之外的概率小于 δ。

WLLN 与 SLLN 的对比

核心区别：

- **收敛模式：**
  - WLLN (依概率收敛 $\xrightarrow{P}$)：

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

    指对于任意小的 ϵ 和 δ，总存在 N，当 $n > N$ 时，$\bar{X}_n$ 落在 $(\mu - \epsilon, \mu + \epsilon)$ 之外的概率小于 δ。
  - SLLN (几乎必然收敛 $\xrightarrow{a.s.}$)：

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

    指 $\bar{X}_n$ 的序列几乎必然会收敛到 μ。也就是说，除了一个概率为零的例外集合，对于每一个样本路径，当 $n \to \infty$ 时，$\bar{X}_n \to \mu$。

WLLN 与 SLLN 的对比

核心区别:

- **收敛模式:**
  - WLLN (依概率收敛 $\xrightarrow{P}$):
    $$\forall \epsilon > 0, \lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$
    指对于任意小的 ϵ 和 δ，总存在 N，当 $n > N$ 时，$\bar{X}_n$ 落在 $(\mu - \epsilon, \mu + \epsilon)$ 之外的概率小于 δ。
  - SLLN (几乎必然收敛 $\xrightarrow{a.s.}$):
    $$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$
    指 $\bar{X}_n$ 的序列几乎必然会收敛到 μ。也就是说，除了一个概率为零的例外集合，对于每一个样本路径，当 $n \to \infty$ 时，$\bar{X}_n \to \mu$。
- **强度:**
  - SLLN 更强。几乎必然收敛 蕴含 依概率收敛 (SLLN $\implies$ WLLN)。

WLLN 与 SLLN 的对比

核心区别:

- **收敛模式:**
  - WLLN (依概率收敛 $\xrightarrow{P}$):
    $$\forall \epsilon > 0, \lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$$
    指对于任意小的 ϵ 和 δ，总存在 N，当 $n > N$ 时，$\bar{X}_n$ 落在 $(\mu - \epsilon, \mu + \epsilon)$ 之外的概率小于 δ。
  - SLLN (几乎必然收敛 $\xrightarrow{a.s.}$):
    $$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$
    指 $\bar{X}_n$ 的序列几乎必然会收敛到 μ。也就是说，除了一个概率为零的例外集合，对于每一个样本路径，当 $n \to \infty$ 时，$\bar{X}_n \to \mu$。
- **强度:**
  - SLLN 更强。几乎必然收敛 蕴含 依概率收敛 (SLLN $\implies$ WLLN)。
- **直观解释:**
  - WLLN：当 n 很大时，$\bar{X}_n$ 不太可能偏离 μ 太远。
  - SLLN：当 n 趋于无穷时，$\bar{X}_n$ 最终会等于 μ (除概率为 0 的例外)。

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**：
  - 为用频率近似概率提供了理论依据 (伯努利大数定律)。
  - 解释了为什么在大量重复试验中，随机事件的平均结果趋于稳定。

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**：
  - 为用频率近似概率提供了理论依据 (伯努利大数定律)。
  - 解释了为什么在大量重复试验中，随机事件的平均结果趋于稳定。
- **统计推断**：
  - 参数估计：用样本均值估计总体期望是矩估计法等方法的基础。
  - 蒙特卡洛方法：通过大量随机抽样和计算样本均值来估计复杂问题的数值解 (如积分、期望)。

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**：
  - 为用频率近似概率提供了理论依据 (伯努利大数定律)。
  - 解释了为什么在大量重复试验中，随机事件的平均结果趋于稳定。
- **统计推断**：
  - 参数估计：用样本均值估计总体期望是矩估计法等方法的基础。
  - 蒙特卡洛方法：通过大量随机抽样和计算样本均值来估计复杂问题的数值解 (如积分、期望)。
- **实际行业应用**：
  - 风险管理与保险：保险公司能够通过大数定律较准确地预测赔付总额，从而制定合理的保费。
  - 质量控制：通过抽样检查产品并计算次品率，来判断整批产品的质量水平。

总结与展望

本章回顾

- **切比雪夫不等式**：提供了概率估计的一个通用（但较宽松）的界。
- **弱大数定律 (WLLN)**：阐述了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：频率依概率收敛于概率。
- **强大数定律 (SLLN)**：阐述了样本均值几乎必然收敛于总体期望 (更强的收敛形式)。
  - 柯尔莫哥洛夫 SLLN：独立同分布、期望存在。
- 大数定律是频率解释概率、参数估计、蒙特卡洛模拟等众多统计思想的理论支柱。

总结与展望

本章回顾

- **切比雪夫不等式**：提供了概率估计的一个通用（但较宽松）的界。
- **弱大数定律 (WLLN)**：阐述了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：频率依概率收敛于概率。
- **强大数定律 (SLLN)**：阐述了样本均值几乎必然收敛于总体期望 (更强的收敛形式)。
  - 柯尔莫哥洛夫 SLLN：独立同分布、期望存在。
- 大数定律是频率解释概率、参数估计、蒙特卡洛模拟等众多统计思想的理论支柱。

展望

大数定律有多种形式和更广泛的推广 (如马尔可夫、辛钦大数定律，以及对非独立序列的推广)。接下来，我们将学习另一个核心的极限定理——中心极限定理，它将告诉我们样本均值围绕总体期望波动的具体分布形态。