

# 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，为许多统计方法的合理性提供了理论基础。
- **重要提示**：大数定律的结论依赖于随机变量满足特定的**条件**（例如，期望存在、方差有限等）。
- 并非所有随机变量序列都满足这些条件。本讲将以一个著名的反例——**柯西分布 (Cauchy distribution)**为例，深入探讨条件的必要性。

# 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  $\text{Var}(X) = \sigma^2$  (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  $\text{Var}(X) = \sigma^2$  (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, 可以提供一個粗略的估计。
- **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, 在特定条件下, 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  $X_1, X_2, \dots, X_n, \dots$  是一列**相互独立**、具有相同期望  $E(X_i) = \mu$  和相同**有限方差**  $\text{Var}(X_i) = \sigma^2 < +\infty$  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:

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

或者等价地:

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

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

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。

# 伯努利弱大数定律

## 频率的稳定性

### 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:

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

伯努利大数定律揭示了“**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: 伯努利大数定律是切比雪夫弱大数定律的一个重要特例, 为用频率估计概率提供了理论依据。



# 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

- 我们看到，切比雪夫弱大数定律要求方差存在且有限。许多其他版本的大数定律则要求更弱但依然关键的条件：**期望的存在性**。
- 那么，是否存在期望不存在的随机变量？**是的，柯西分布就是一个典型例子。**

# 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

- 我们看到，切比雪夫大数定律要求方差存在且有限。许多其他版本的大数定律则要求更弱但依然关键的条件：**期望的存在性**。
- 那么，是否存在期望不存在的随机变量？**是的，柯西分布就是一个典型例子。**

### 柯西分布 (Cauchy Distribution)

一个更普适的柯西分布的概率密度函数 (PDF) 定义为：

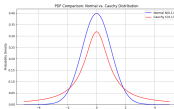
$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[ 1 + \left( \frac{x-x_0}{\gamma} \right)^2 \right]}$$

其中  $x_0 \in (-\infty, +\infty)$  是位置参数 (决定峰值位置，也是中位数和众数)， $\gamma > 0$  是尺度参数 (决定分布的宽度，半峰全宽)。

当  $x_0 = 0, \gamma = 1$  时，即为标准柯西分布：

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{特点:}$$

- 形状类似钟形 (对称于  $x_0$ )，但尾部比“正态分布”厚得多 (重尾)。
- **期望  $E(X)$  不存在**：对于任意  $x_0, \gamma$ ，其期望积分  $\int_{-\infty}^{\infty} xf(x; x_0, \gamma) dx$  仍然是发散的 (更准确地说，绝对期望发散)。
- **方差  $\text{Var}(X)$  也不存在** (因为它依赖于期望)。



图：\*

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

# 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

- 我们看到，切比雪夫弱大数定律要求方差存在且有限。许多其他版本的大数定律则要求更弱但依然关键的条件：**期望的存在性**。
- 那么，是否存在期望不存在的随机变量？**是的，柯西分布就是一个典型例子。**

### 柯西分布 (Cauchy Distribution)

一个更普适的柯西分布的概率密度函数 (PDF) 定义为：

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

其中  $x_0 \in (-\infty, +\infty)$  是位置参数 (决定峰值位置，也是中位数和众数)， $\gamma > 0$  是尺度参数 (决定分布的宽度，半峰全宽)。

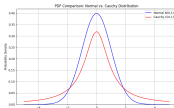
当  $x_0 = 0, \gamma = 1$  时，即为标准柯西分布：

$$f(x) = \frac{1}{\pi(1+x^2)} \quad \text{特点:}$$

- 形状类似钟形 (对称于  $x_0$ )，但尾部比正态分布“厚得多” (重尾)。
- **期望  $E(X)$  不存在**：对于任意  $x_0, \gamma$ ，其期望积分  $\int_{-\infty}^{\infty} xf(x; x_0, \gamma) dx$  仍然是发散的 (更准确地说，绝对期望发散)。
- **方差  $\text{Var}(X)$  也不存在** (因为它依赖于期望)。

这意味着什么？

- 柯西分布不满足绝大多数大数定律版本所要求的“期望存在”或“方差有限”的条件。
- 我们将看到，对于 i.i.d. 柯西分布的随机变量序列，它们的样本均值**不会**收敛到任何常数。



图：\*

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

# 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  $N(0, 1)$   
( $E(X) = 0, \text{Var}(X) = 1$ )

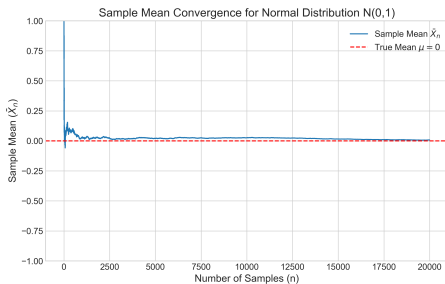


图:  $\bar{X}_n$  快速稳定收敛到  $\mu = 0$

柯西分布  $C(0, 1)$   
( $E(X)$  不存在)

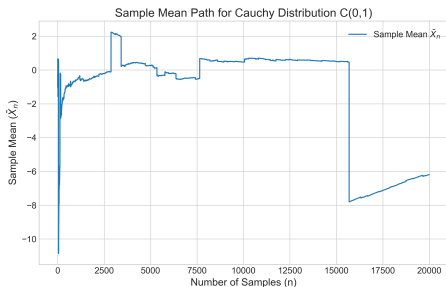


图:  $\bar{X}_n$  持续剧烈波动, 不收敛

## 观察与结论：

- 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。

## 7.2 弱大数定律 (续): 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)

辛钦弱大数定律对随机变量序列的方差要求更低, 仅要求期望存在且序列是独立同分布的。

### 定理 7.4 (辛钦弱大数定律)

设  $X_1, X_2, \dots, X_n, \dots$  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  $E(X_i) = \mu$  **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:

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

即  $\bar{X}_n \xrightarrow{P} \mu$ 。

回顾柯西分布:

- 柯西分布是 i.i.d. 序列, 但其期望**不存在**。
- 因此, i.i.d. 柯西随机变量序列**不满足**辛钦弱大数定律的条件。
- 其样本均值  $\bar{X}_n$  **不会**依概率收敛到任何常数  $\mu$ 。这与定律的结论相反。

## 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)

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, 同样要求期望存在。

### 定理 7.5 (柯尔莫哥洛夫强大数定律)

设  $X_1, X_2, \dots, X_n, \dots$  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  $E(X_i) = \mu$  **存在** (即  $|\mu| < \infty$ )。令样本均值为  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ 。则:

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

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

再次回顾柯西分布:

- 柯西分布的期望**不存在**。
- 因此, i.i.d. 柯西随机变量序列**不满足**柯尔莫哥洛夫强大数定律的条件。
- 其样本均值  $\bar{X}_n$  **不会**几乎必然收敛到任何常数。

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。

# WLLN 与 SLLN 的区别与联系 (1): 核心概念与正态分布

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 ( $\xrightarrow{P}$ ):  $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$ .  
描述的是概率趋于 1。
- 强大数定律 (SLLN) - 几乎必然收敛 ( $\xrightarrow{a.s.}$ ):  $P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$ .  
描述的是样本路径的极限行为。

# WLLN 与 SLLN 的区别与联系 (1): 核心概念与正态分布

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 ( $\xrightarrow{P}$ ):  $\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| \geq \epsilon) = 0$ .  
描述的是概率趋于 1。
- 强大数定律 (SLLN) - 几乎必然收敛 ( $\xrightarrow{a.s.}$ ):  $P(\lim_{n \rightarrow \infty} \bar{X}_n = \mu) = 1$ .  
描述的是样本路径的极限行为。

条件与对比:

- SLLN 比 WLLN 更强 ( $\xrightarrow{a.s.} \implies \xrightarrow{P}$ )。
- 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  $\mu$  存在, 方差  $\sigma^2$  有限。满足大数定律条件。
  - i.i.d. 正态随机变量的样本均值  $\bar{X}_n$  依概率收敛 且 几乎必然收敛 于其期望  $\mu$ 。



# 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

回顾柯西分布的关键特性:

- 期望不存在。因此不满足大数定律的期望存在条件。

# 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

回顾柯西分布的关键特性:

- 期望**不存在**。因此不满足大数定律的期望存在条件。

对于柯西分布的样本均值:

- i.i.d. 柯西随机变量的样本均值  $\bar{X}_n$  **不**依概率收敛于任何常数。
- i.i.d. 柯西随机变量的样本均值  $\bar{X}_n$  **不**几乎必然收敛于任何常数。
- **惊人事实**: 若  $X_1, \dots, X_n$  i.i.d. 服从标准柯西分布, 则它们的样本均值  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$  **仍服从标准柯西分布!** (与  $n$  无关)

这与满足大数定律条件的正态分布形成了鲜明对比, 突出了期望存在性对于大数定律的重要性。

## 7.4 大数定律的应用 (Applications of LLN)

应用的基础：满足定律的条件

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：

- **前提：**这些应用都隐式或显式地假设了相关的随机变量序列满足大数定律的条件 (通常是期望存在)。
- **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。**但只有当总体期望存在时才成立！**
  - **蒙特卡洛方法：**通过大量随机抽样来估计积分、期望等数值解。依赖于“样本均值逼近期望”的性质。
- **风险管理与保险：**预测平均赔付，依赖于平均结果的稳定性。
- **质量控制与抽样检验：**根据样本均值推断总体均值或合格率，依赖于样本均值的收敛性。
- **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。
  - 如果测量误差服从柯西分布，那么增加测量次数并计算平均值是**没有帮助的**，因为平均值不会收敛到真实值，甚至其分布形状都不变！

大数定律的条件 (尤其是期望存在) 是其应用有效性的关键。对于像柯西分布这样期望不存在的情况，基于样本均值的大数定律应用会失效。

## 本讲小结

- 切比雪夫不等式：提供了概率的通用上界（需方差存在）。
- 大数定律：样本均值在特定条件下收敛于总体期望。
  - 弱大数定律 (WLLN)：依概率收敛 ( $\xrightarrow{P}$ )。辛钦 WLLN 要求 i.i.d. 且期望存在。
  - 强大数定律 (SLLN)：几乎必然收敛 ( $\xrightarrow{a.s.}$ )。柯尔莫哥洛夫 SLLN 要求 i.i.d. 且期望存在。结论更强。
- **条件的重要性**：大数定律的结论依赖于随机变量满足特定条件，最核心的是**期望的存在性**。
- **反例：柯西分布**
  - 柯西分布的期望**不存在**。
  - 因此，i.i.d. 柯西变量的样本均值**不服从**大数定律，不会收敛到任何常数。
  - 柯西分布的样本均值**仍服从**柯西分布，突显了期望不存在时的特殊行为。
- 大数定律是连接概率论与统计实践的**核心桥梁**，理解其条件和局限性对于正确应用至关重要。

## 展望

- 大数定律描述了平均行为的极限。
- 另一类核心极限定理是中心极限定理 (CLT)，它描述了样本均值（或和）的分布趋向于正态分布。它也通常要求期望和方差存在。
- 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！

问题与讨论