



中国科学院数学与系统科学研究院

Academy of Mathematics and Systems Science
Chinese Academy of Sciences

第七章 收敛和极限定理

洪永淼

中国科学院数学与系统科学研究院

中国科学院大学经济与管理学院

Copyright © 2024 by Professor Hong Yongmiao, All rights reserved. Requests for permission should be mailed to: ymhong@amss.ac.cn

1. 版权归作者洪永淼教授所有；
2. 不得移除作者署名，否则将视为侵权；
3. 对于不遵守此声明或者其他违法使用本文内容者，作者依法保留追究权等。
4. 发现课件错误请联系作者 ymhong@amss.ac.cn

◆ 问题

如何判断样本均值 \bar{X}_n 是否是均值 μ 的一个好的估计量?

- 利用**均方误差** (Mean Square Error, MSE), 在 IID 假设下,

$$\begin{aligned} \text{MSE}(\bar{X}_n) &= E(\bar{X}_n - \mu)^2 \\ &= \frac{\sigma^2}{n} \rightarrow 0 \text{ as } n \rightarrow \infty \end{aligned}$$

◆ 问题

除了均方误差 (MSE) 外, 是否可以利用其它的测度评判估计量?

- 有, 本章将介绍该内容。

◆ 问题

若 X^n 满足 IID 正态假设, 有

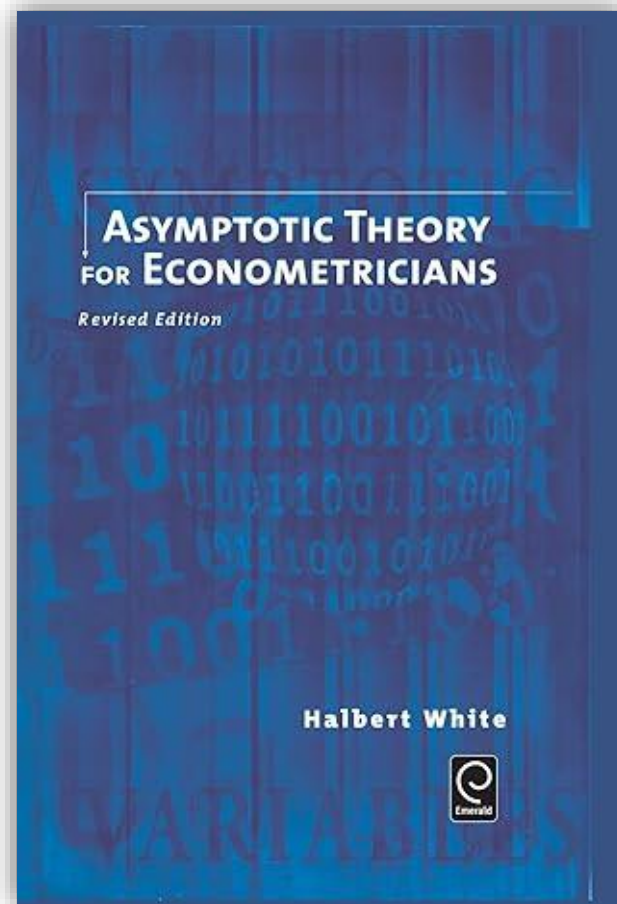
$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1), \text{ 对所有 } n \geq 1.$$

若 X_i 不再满足正态分布, Z_n 是否还是正态分布?

- 当 $n \rightarrow \infty$ 时, 利用**渐近理论 (Asymptotic theory)** 或**大样本理论 (Large sample theory)**。

参考书推荐

- White H. Asymptotic theory for econometricians (Revised Edition). Academic press, 2014.



第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定义 7.1 [极限 (Limit)]

- 令 $\{b_n, n = 1, 2, \dots\}$ 为一个非随机实数序列。
- 若存在实数 b , 使得对每一给定实数 $\epsilon > 0$, 均存在有限整数 $N(\epsilon)$ 满足对所有 $n \geq N(\epsilon)$, 有 $|b_n - b| < \epsilon$, 则 b 是序列 $\{b_n, n = 1, 2, \dots\}$ 的**极限**。

- 序列 b_n 的极限可写作:

$$b = \lim_{n \rightarrow \infty} b_n$$

- ϵ 可解释为预先设定的对 b_n 和 b 之间**偏离** (deviation) 的**容忍度** (tolerance level)。

例 7.1:

- 令 $b_n = 1 - n^{-1}$ 。则当 $n \rightarrow \infty$ 时, $b_n \rightarrow 1$ 。
 - ✓ 因为对 $b = 1$ 和任意 $\epsilon > 0$, 存在 $N(\epsilon) = [\epsilon^{-1}] + 1$, 其中 $[\cdot]$ 表示整数部分, 使得对所有 $n \geq N(\epsilon)$, 有

$$|b_n - b| = \frac{1}{n} < \epsilon$$

- 本例中,
 - ✓ 若 $\epsilon = 10^{-4}$, 则需 $N(\epsilon) = 1/\epsilon + 1 = 10^4 + 1$;
 - ✓ 若 $\epsilon = 10^{-8}$, 则需 $N(\epsilon) = 10^8 + 1$ 。

例 7.2:

- 令 $b_n = \left(1 + \frac{a}{n}\right)^n$, 其中 a 为常数。则当 $n \rightarrow \infty$ 时, $b_n \rightarrow e^a$ 。

例 7.3:

- 令 $b_n = (-1)^n$ 。对常数 $M > 1$ 和所有 $n \geq 1$, 有 $|b_n| \leq M$, 故 $\{b_n, n = 1, 2, \dots\}$ 有界, 但其极限不存在。

解:

- 令 $\epsilon = 1/2$ 。则不存在一个常数 b 和一个整数 $N(\epsilon)$ 使得对所有 $n > N(\epsilon)$, 有 $|b_n - b| < \epsilon$ 。

定义 7.2 [连续性 (Continuity)]:

若对任意实数序列 $\{b_n, n = 1, 2, \dots\}$, 当 $n \rightarrow \infty$, $b_n \rightarrow b$ 时, 有 $g(b_n) \rightarrow g(b)$, 则称函数 $g: \mathbb{R} \rightarrow \mathbb{R}$ 在 b 点处连续。

- **另一个等价定义:** 对任意给定的常数 $\epsilon > 0$, 存在 $\delta = \delta(\epsilon)$ 使得对任意 $|b_n - b| < \delta$, 有 $|g(b_n) - g(b)| < \epsilon$ 。
- **连续函数序列的极限等于函数在极限处的值:** 当 $g(\cdot)$ 在 b 处连续时, 可记作

$$\lim_{n \rightarrow \infty} g(b_n) = g(\lim_{n \rightarrow \infty} b_n) = g(b)$$

例 7.4:

- 假设 $n \rightarrow \infty$ 时, $a_n \rightarrow a$ 且 $b_n \rightarrow b$ 。
- 则当 $n \rightarrow \infty$ 时,
 - (1) $a_n + b_n \rightarrow a + b$;
 - (2) $a_n b_n \rightarrow ab$;
 - (3) 若 $b \neq 0$, $a_n/b_n \rightarrow a/b$ 。

例 7.5:

- 定义函数

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{1}{2}, & 0 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

- 这里, $F(x)$ 在 0 和 1 点处均不连续。

✓ $F(x)$ 在 0 处不连续: 因为至少存在一个序列 $\{b_n\}$, 如

$b_n = -\frac{1}{n}$, 有 $b_n \rightarrow b = 0$ 。但对所有 $n \geq 1$, $F(b_n) =$

$F(-\frac{1}{n}) = 0$ 而 $\lim_{n \rightarrow \infty} F(-\frac{1}{n}) = 0 \neq F(0) = 1/2$ 。

✓ 试找到一个序列 $\{b_n\}$ 验证 $F(x)$ 在 1 处不连续?

定义 7.3 [数量级 (Order of Magnitude)]:

(1) 若对**某个**足够大的实数 $M < \infty$, 存在一个有限整数 $N(M)$ 使得

$$\text{对所有 } n \geq N(M), \text{ 有 } |n^{-\lambda} b_n| < M,$$

则序列 $\{b_n, n = 1, 2, \dots\}$ 的**最高数量级**为 n^λ , 记作 $b_n = O(n^\lambda)$ 或 $n^{-\lambda} b_n = O(1)$;

(2) 若对**每个**实数 $\epsilon > 0$, 都存在一个有限整数 $N(\epsilon)$ 使得

$$\text{对所有 } n \geq N(\epsilon), \text{ 有 } |n^{-\lambda} b_n| < \epsilon,$$

则序列 $\{b_n\}$ 的**数量级小于** n^λ , 记作 $b_n = o(n^\lambda)$ 或 $n^{-\lambda} b_n = o(1)$ 。

例 7.6:

- $b_n = 4 + 2n + 6n^2$, 则 $b_n = O(n^2)$, 因为对所有充分大的 n , 有

$$\frac{b_n}{n^2} = \frac{4}{n^2} + \frac{2n}{n^2} + \frac{6n^2}{n^2} \rightarrow 6 < 2M = 2 \times 6 \text{ (当 } M = 6 \text{ 时)}$$

- 直观上, b_n 的数量级由最快趋向无穷的那一项主导, 即 n^2 决定。

若 $b_n = O(n^\lambda)$, 则 $\left|\frac{b_n}{n^\lambda}\right|$ 有界

- 有可能 $\lim_{n \rightarrow \infty} \frac{b_n}{n^\lambda}$ 不存在但 $\left|\frac{b_n}{n^\lambda}\right|$ 有界, 此时仍有 $b_n = O(n^\lambda)$ 。

例 7.7:

- 令 $b_n = (-1)^n$, 则 $b_n = O(1)$, 因为对所有 $n \geq 1$,

$$|b_n| = 1 < M \equiv 1.01$$

若 $b_n = o(n^\lambda)$, 则 $\lim_{n \rightarrow \infty} \frac{b_n}{n^\lambda} = 0$

- 在 $b_n = o(n^\lambda)$ 的定义中, 常数 ϵ 可取很小的值。 ϵ 越小, $N(\epsilon)$ 越大。直观上, $b_n = o(n^\lambda)$ 表示 b_n 的增速严格小于 n^λ , 即

$$\lim_{n \rightarrow \infty} \frac{b_n}{n^\lambda} = 0$$

- 但 $b_n = O(n^\lambda)$ 时, $\lim_{n \rightarrow \infty} \frac{b_n}{n^\lambda}$ 可能不存在。

例 7.8:

- 令 $b_n = 4 + 2n + 6n^2$, 则对所有 $\delta > 0$, $b_n = o(n^{2+\delta})$ 。

若 $b_n = o(n^\lambda)$, 则 $b_n = O(n^\lambda)$

- 直观上, 若 b_n 以低于 n^λ 的速度增加, 当然其最大速率不超过 n^λ 。

引理 7.1

• 令 a_n 和 b_n 为标量。

(1) 若 $a_n = O(n^\lambda)$ 和 $b_n = O(n^\mu)$,

则 $a_n b_n = O(n^{\lambda+\mu})$, $a_n + b_n = O(n^k)$, 其中 $k = \max(\lambda, \mu)$;

(2) 若 $a_n = o(n^\lambda)$ 和 $b_n = o(n^\mu)$,

则 $a_n b_n = o(n^{\lambda+\mu})$, $a_n + b_n = o(n^k)$, 其中 $k = \max(\lambda, \mu)$;

(3) 若 $a_n = O(n^\lambda)$ 和 $b_n = o(n^\mu)$,

则 $a_n b_n = o(n^{\lambda+\mu})$, $a_n + b_n = O(n^k)$, 其中 $k = \max(\lambda, \mu)$ 。

证明:

(1)

- 因为 a_n 的速率不超过 n^λ , b_n 的速率不超过 n^μ , 二者乘积 $a_n b_n$ 的速率将不超过 $n^{\lambda+\mu}$ 。这是因为对所有充分大的 n , 即对所有 $n \geq N(M)$, 有

$$\left| \frac{a_n b_n}{n^{\lambda+\mu}} \right| = \left| \frac{a_n}{n^\lambda} \frac{b_n}{n^\mu} \right| \leq M \times M = M^2$$

- 另一方面, $a_n + b_n$ 将由更快趋向无穷的那一项主导: 对所有充分大的 n , 有

$$\left| \frac{a_n + b_n}{n^k} \right| = \left| \frac{a_n}{n^k} + \frac{b_n}{n^k} \right| \leq \epsilon + M \leq 2M$$

因此, $a_n + b_n = O(n^k)$ 。

证明 (Cont.):

- (2) 与结果 (1) 的证明类似。
- (3) 乘积 $a_n b_n = o(n^{\lambda+\mu})$, 因为给定 $a_n = O(n^\lambda)$ 和 $b_n = o(n^\mu)$, 当 $n \rightarrow \infty$ 时, 有

$$\left| \frac{a_n b_n}{n^{\lambda+\mu}} \right| = \left| \frac{a_n}{n^\lambda} \frac{b_n}{n^\mu} \right| \leq M \left| \frac{b_n}{n^\mu} \right| \rightarrow 0$$

证毕。

例 7.9:

设 $a_n = O(1)$, $b_n = o(1)$, 则 $a_n b_n = o(1)$ 和 $a_n + b_n = O(1)$ 。

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

◆ 问题

为什么统计学和计量经济学需要**收敛 (convergence)** 概念呢?

回顾

- 样本容量为 n 的**随机样本** $X^n = (X_1, \dots, X_n)$ 是由 n 个随机变量 X_1, \dots, X_n 组成的一个序列，它可视为一个 n 维实值随机向量，其中维数 n 可趋于无穷大。
- X^n 的**实现值 (realization)** 是一个 n 维向量 $x^n = (x_1, \dots, x_n)$ ，通常称为从随机样本 X^n 生成的一个**样本点 (sample point)** 或**数据集 (data set)**。

回顾 (Cont.)

- 当 \mathbf{X}^n 是来自总体 PMF/PDF $f_X(\cdot)$ 的 IID 随机样本时, \mathbf{X}^n 的 **联合 PMF/PDF** 为

$$f_{\mathbf{X}^n}(\mathbf{x}^n) = \prod_{i=1}^n f_X(x_i)$$

- 该联合概率分布称为随机样本 \mathbf{X}^n 的**抽样分布**, 它完全描述了随机样本 \mathbf{X}^n 的概率分布。

回顾 (Cont.)

- **参数建模方法**: 假定 $f_X(x)$ 为参数模型, 即存在某有限维数的参数值 θ 使 $f_X(x) = f(x, \theta)$, 其中函数形式 $f(\cdot, \cdot)$ 已知但参数值 θ 未知。

✓ 例如, 若假设 $f_X(x)$ 服从正态分布 $N(\mu, \sigma^2)$, 则有

$$\begin{aligned} f_X(x) &= f(x, \theta) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty \end{aligned}$$

其中 $\theta = (\mu, \sigma^2)$ 。

回顾 (Cont.)

- **统计分析的一个重要目的是：**用给定随机样本 X^n 所生成的数据集 x^n 估计未知参数 θ 。
 - ✓ 参数 θ 的估计量是 X^n 的函数，因此它是一个统计量。
- 需要强调，统计量 $Z_n = T(X^n)$ 仅为 X^n 的函数，而不包含任何未知参数，它本身是随机变量或随机向量。

回顾 (Cont.)

- 为说明各种收敛概念的重要性，现在考察两个简单的统计量——**样本均值**和**样本方差**。
- 令总体 X^n 为一个来自**均值为 μ** ，**方差为 σ^2** 的总体分布的 IID 随机样本，其样本容量是 n 。

✓ **样本均值估计量**

$$\bar{X}_n = T(X^n) = \frac{1}{n} \sum_{i=1}^n X_i$$

✓ **样本方差估计量**

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$$

◆ 问题 7.1

如何测量 \bar{X}_n 对 μ 的接近程度和 S_n^2 对 σ^2 的接近程度呢？

- **[估计量的随机性]**: 由于 \bar{X}_n 和 S_n^2 为随机变量, 因此第七章第一节有关非随机序列的收敛概念不能应用。
- \bar{X}_n 和 S_n^2 均是从 S 到实数集的映射, 其中 S 是随机试验的样本空间:
 - ✓ $\bar{X}_n : S \rightarrow \mathbb{R}$,
 - ✓ $S_n^2 : S \rightarrow \mathbb{R}^+$ 。

CONTINUE

- 假设进行了一项随机试验，出现某个**基本结果** $s \in S$ ，并观测到一个数据集 $\mathbf{x}^n = (x_1, \dots, x_n)$ ，其中 $x_i = X_i(s)$ ，数据集 \mathbf{x}^n 为随机样本 \mathbf{X}^n 的一个实现。从数据集 \mathbf{x}^n 可计算 μ 的一个估计值 $\bar{x}_n = \bar{X}_n(s)$ 和 σ^2 的一个估计值 $s_n^2 = S_n^2(s)$ 。
- 事实上，每个基本结果 $s \in S$ 都将生成实数序列 $\{\bar{x}_n = \bar{X}_n(s), n = 1, 2, \dots\}$ 和 $\{s_n^2 = S_n^2(s), n = 1, 2, \dots\}$ 。这两个非随机序列分别称为基本结果 s 发生时的样本均值 \bar{X}_n 和样本方差 S_n^2 的**样本路径** (sample path)。

CONTINUE

- \bar{X}_n 和 S_n^2 均有许多此类样本路径，不同样本路径对应不同的基本结果 $s \in S$ ，因此需要定义适当的收敛概念以测度 \bar{X}_n 和 μ 之间以及 S_n^2 和 σ^2 之间的距离。
- 存在多种收敛概念，这些不同收敛概念的**共同特征**就是它们均定义**绝大多数非随机序列** $\{\bar{X}_n(s), n = 1, 2, \dots\}$ 和 $\{S_n^2(s), n = 1, 2, \dots\}$ 分别**收敛于** μ 和 σ^2 。

◆ 问题

绝大多数非随机序列 $\{\bar{X}_n(s), n = 1, 2, \dots\}$ 和 $\{S_n^2(s), n = 1, 2, \dots\}$ 指的是什么？

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定义 7.4 [依二次方均值收敛 (Convergence in Quadratic Mean)]

- 令 $\{Z_n, n = 1, 2, \dots\}$ 为一个随机变量序列, Z 为随机变量。若当 $n \rightarrow \infty$ 时, 有

$$E(Z_n - Z)^2 \rightarrow 0$$

或等价地

$$\lim_{n \rightarrow \infty} E(Z_n - Z)^2 = 0$$

则称随机序列 $\{Z_n, n = 1, 2, \dots\}$ 依二次方均值 (或依均方(mean square)) 收敛于 Z , 也可记作

$$Z_n \xrightarrow{q.m.} Z$$

或

$$Z_n - Z = o_{q.m.}(1)。$$

定义 7.4 的含义

- 依二次方均值收敛表示当 $n \rightarrow \infty$ 时, $Z_n(s)$ 和 $Z(s)$ 之间偏离 (deviation) 平方的权重平均值趋于 0。
- 此处是对样本空间 S 所有的基本结果 $\{s\}$ 按照其发生的概率赋予权重并求平均值。

$$E(Z_n - Z)^2 = \int_{-\infty}^{\infty} [Z_n(s) - Z(s)]^2 dF(s)$$

其中 $F(s)$ 为基本结果 s 的 CDF。

- 当 Z_n 依二次方均值收敛于 Z 时, 可能在某些样本路径上 $Z_n(s)$ 并不收敛于 $Z(s)$ 甚至发散。然而, 当 n 不断增大时, 这些样本路径的二次方偏离加权后可忽略不计。

例 7.10:

- 假设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为来自均值为 μ , 方差为 σ^2 总体分布的 IID 随机样本。定义 $Z_n = \bar{X}_n$ 。证明 $\bar{X}_n \xrightarrow{q.m.} \mu$ 。

证明:

- 只需证明 $\lim_{n \rightarrow \infty} E(\bar{X}_n - \mu)^2 = 0$ 。
- 因为 $E(\bar{X}_n) = \mu$ 且 $\text{var}(\bar{X}_n) = \frac{\sigma^2}{n}$, 当 $n \rightarrow \infty$ 有

$$\begin{aligned} E(\bar{X}_n - \mu)^2 &= \text{var}(\bar{X}_n) \\ &= \frac{\sigma^2}{n} \\ &\rightarrow 0 \end{aligned}$$

定义 7.5 [L_p -收敛 (L_p -Convergence)]

- 假设 $0 < p < \infty$, $\{Z_n, n = 1, 2, \dots\}$ 为满足 $E|Z_n|^p < \infty$ 的一个随机变量序列, 而 Z 为满足 $E|Z|^p < \infty$ 的一个随机变量。若

$$\lim_{n \rightarrow \infty} E|Z_n - Z|^p = 0$$

则随机变量序列 $\{Z_n\}$ 依 L_p 范数收敛于 Z 。

◆ 问题

上述分析中 Z_n 和 Z 均为标量。若 Z_n 和 Z 为 d 维随机向量, 其中正整数 d 固定 (即当 $n \rightarrow \infty$ 时, d 不变), 那么如何定义 L_p -收敛呢?

- 对 $i = 1, \dots, d$, 若向量 Z_n 的每一个元素都有 $Z_{in} \xrightarrow{L_p} Z_i$, 则随机向量序列 $\{Z_n, n = 1, 2, \dots\}$ 依 L_p 收敛于 Z 。
- 换言之, 按元素逐个收敛保证整个向量 Z_n 联合收敛, 反之亦然。

目 录

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定义 7.6 [依概率收敛 (Convergence in Probability)]

若对每个常数 $\epsilon > 0$, 当 $n \rightarrow \infty$ 时, 有

$$P[|Z_n - Z| > \epsilon] \rightarrow 0$$

则称随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 依概率收敛于随机变量 Z 。

- 当 Z_n 依概率收敛于 Z 时, 对所有 $\epsilon > 0$, 有

$$\lim_{n \rightarrow \infty} P(|Z_n - Z| > \epsilon) = 0$$

- 或记作

$$p \lim_{n \rightarrow \infty} Z_n = Z, \quad Z_n \xrightarrow{p} Z, \quad Z_n - Z = o_p(1), \quad Z_n - Z \xrightarrow{p} 0$$

依概率收敛的含义

- 依概率收敛因其较易检验，在统计学和计量经济学中应用广泛。依概率收敛也称**弱收敛** (weak convergence)。
- 定义样本空间 S 的一个集合

$$A_n(\epsilon) = \{s \in S : |Z_n(s) - Z(s)| \leq \epsilon\}$$

即 $A_n(\epsilon)$ 是样本空间 S 的一个子集，其中所有基本结果 $s \in S$ ，均满足 $|Z_n(s) - Z(s)| \leq \epsilon$ 。 $A_n(\epsilon)$ 的大小同时取决于 ϵ 与 n 。当 $n \rightarrow \infty$ ， Z_n 依概率收敛于 Z 时，发生“**小偏离**”的概率为 $P[|Z_n - Z| \leq \epsilon] = P[A_n(\epsilon)] \rightarrow 1$ 对任意有限的 n ， $P[A_n(\epsilon)]$ 可能不等于 1，但当 $n \rightarrow \infty$ 时其将任意接近于 1。

依概率收敛的含义 (Cont.)

- 换言之, 补集 $A_n^c(\epsilon)$ 表示样本空间 S 中出现“**大偏离**”事件, 这个集合对所有有限 n 都可能存在。然而, 当 $n \rightarrow \infty$ 时, 其出现的概率将衰减至 0。因此, 依概率收敛也称为**依概率趋近 1 收敛** (convergence with probability approaching 1)。

例 7.11:

- 假设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为来自均匀分布 $U[0, \theta]$ 的 IID 随机样本, 其中 $\theta > 0$ 是未知参数。定义统计量 $Z_n = \max_{1 \leq i \leq n} X_i$ 。
- 问 Z_n 是 θ 的一致估计量吗?

解:

- 因为 $\{|Z_n - \theta| > \epsilon\} = \{Z_n - \theta > \epsilon\} \cup \{Z_n - \theta < -\epsilon\}$

- 对任意给定 $\epsilon > 0$, 当 $n \rightarrow \infty$ 时, 有

$$\begin{aligned} P(|Z_n - \theta| > \epsilon) &= P(Z_n > \theta + \epsilon) + P(Z_n < \theta - \epsilon) = P(Z_n < \theta - \epsilon) \\ &= P\left[\max_{1 \leq i \leq n} (X_i) < \theta - \epsilon\right] \\ &= P(X_1 < \theta - \epsilon, X_2 < \theta - \epsilon, \dots, X_n < \theta - \epsilon) \\ &= \prod_{i=1}^n P(X_i < \theta - \epsilon) \quad (\text{根据独立性}) \\ &= \left(\frac{\theta - \epsilon}{\theta}\right)^n = \left(1 - \frac{\epsilon}{\theta}\right)^n \rightarrow 0 \end{aligned}$$

- 因此 Z_n 是 θ 的一致估计量。

- **注:** 统计量 $Z_n = \max_{1 \leq i \leq n} X_i$ 称为**次序统计量** (order statistic), 其

包含了对随机样本 X^n 中的 n 个随机变量的某种排序信息。

- **样本中位数** (sample median) 也是次序统计量。

数量级为 n^α 的依概率收敛, 其中 α 可为正数或负数:

- (1) 若当 $n \rightarrow \infty$ 时, $Z_n/n^\alpha \xrightarrow{p} 0$, 则称随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 为**依概率数量级小于** n^α , 记作

$$Z_n = o_p(n^\alpha)$$

- (2) 若对任意给定 $\delta > 0$, 存在常数 $M = M(\delta) < \infty$ 和有限整数 $N = N(\delta)$, 使得对所有 $n > N$, 有

$$P(|Z_n/n^\alpha| > M) < \delta$$

则称随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 为**依概率数量级不超过** n^α , 记作

$$Z_n = O_p(n^\alpha)$$

数量级为 n^α 的依概率收敛，其中 α 可为正数或负数 (Cont.):

- 若 $Z_n = O_p(n^\alpha)$,
 - ✓ 其中 $\alpha > 0$, 则数量级 n^α 是 Z_n 以趋近 1 的概率趋向无穷大的最快发散速率。
 - ✓ 当 $\alpha < 0$ 时, 数量级 n^α 是 Z_n 以趋近 1 的概率趋于 0 的最慢收敛速率。
- 事实上, $Z_n = O_p(n^\alpha)$ 的定义涉及以下依概率有界的概念。

定义 7.7 [依概率有界 (Boundedness in Probability)]

对任意常数 $\delta > 0$, 存在常数 $M = M(\delta)$ 和有限整数 $N = N(\delta)$, 使得对所有 $n \geq N$, 有

$$P(|Z_n| > M) < \delta,$$

则称 Z_n 依概率有界, 记作

$$Z_n = O_p(1).$$

例 7.12:

若对所有 $n \geq 1$, 有 $Z_n \sim N(0, 1)$, 则 $Z_n = O_p(1)$ 。

解:

- 对任意给定的 $\delta > 0$, 存在有限常数 $M = \Phi^{-1}(1 - \delta/2) < \infty$, 其中 $\Phi^{-1}(\cdot)$ 为 $N(0, 1)$ 的 CDF $\Phi(\cdot)$ 的反函数, 满足对所有 $n \geq 1$ 有

$$P(|Z_n| > M) = 2[1 - \Phi(M)] = \delta < 2\delta$$

引理 7.2 [马尔可夫 (Markov) 不等式]:

假设 X 为随机变量, $g(X)$ 为非负函数。则对任意 $\epsilon > 0$ 和任意 $k > 0$, 有

$$P[g(X) \geq \epsilon] \leq \frac{E[g(X)^k]}{\epsilon^k}$$

证明:

- 令 $\mathbf{1}(\cdot)$ 为指示函数, 取值为 0 或 1。
- 则

$$\begin{aligned}
 P[g(X) > \epsilon] &= \int_{\{x:g(x)>\epsilon\}} dF_X(x) = \int_{-\infty}^{\infty} \mathbf{1}[g(x) > \epsilon] dF_X(x) \\
 &\leq \int_{-\infty}^{\infty} \mathbf{1}[g(x) > \epsilon] \frac{g(x)^k}{\epsilon^k} dF_X(x) \leq \int_{-\infty}^{\infty} \frac{g(x)^k}{\epsilon^k} dF_X(x) \\
 &= \frac{1}{\epsilon^k} E[g(X)^k]
 \end{aligned}$$

证毕。

马尔可夫不等式的重要意义

- 马尔可夫不等式是证明依概率收敛的标准工具，其用矩约束条件界定了尾部概率。尾部概率的厚度依赖于概率分布的矩。分布的矩的值越大，分布的尾部概率越大。
- 当 $k = 2$ 和 $g(x) = |x|$ 时，马尔可夫不等式称为**切比雪夫 (Chebyshev) 不等式**。由于二阶矩的计算较为简单，因此切比雪夫不等式在统计学和计量经济学应用广泛。

定理 7.1 [弱大数定律 (Weak Law of Large Numbers, WLLN)]

假设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为来自均值 $E(X_i) = \mu$ 和方差 $\text{var}(X_i) = \sigma^2 < \infty$ 的总体分布的 IID 随机样本。定义样本均值 $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ ，则对任意给定常数 $\epsilon > 0$ ，当 $n \rightarrow \infty$ 时，有

$$P[|\bar{X}_n - \mu| \leq \epsilon] \rightarrow 1$$

或

$$\bar{X}_n - \mu \xrightarrow{p} 0,$$

或

$$\bar{X}_n - \mu = o_p(1)$$

证明:

- 首先, 注意到 $E(\bar{X}_n) = \mu$ 和 $\text{var}(\bar{X}_n) = \sigma^2/n$ 。
- 由切比雪夫不等式 ($g(\bar{X}_n) = |\bar{X}_n - \mu|$ 和 $k = 2$) , 有

$$\begin{aligned} P[|\bar{X}_n - \mu| > \varepsilon] &\leq \frac{E(\bar{X}_n - \mu)^2}{\varepsilon^2} \\ &= \frac{\text{var}(\bar{X}_n)}{\varepsilon^2} \\ &= \frac{\sigma^2}{n\varepsilon^2} \end{aligned}$$

- 则当 $n \rightarrow \infty$ 时,

$$\begin{aligned} P[|\bar{X}_n - \mu| \leq \varepsilon] &= 1 - P[|\bar{X}_n - \mu| > \varepsilon] \\ &\geq 1 - \frac{\sigma^2}{n\varepsilon^2} \\ &\rightarrow 1 \end{aligned}$$

- 因此, $\bar{X}_n \xrightarrow{p} \mu$ 或 $\bar{X}_n - \mu = o_p(1)$ 。 **证毕。**

WLLN 的有限方差假设

- 上述 WLLN 假设了**有限方差**。
- 这个假设在大多数应用中是合理的，但它比所需要的假设要严格。事实上，只需要矩条件满足 $E|X_i| < \infty$ 即可（参见 Resnik, 1999 或 Billingsley, 1995）。
- 为了说明为何 WLLN 需要一定的矩条件，可考虑来自柯西分布的 IID 随机样本的例子，其各阶矩均不存在。

例 7.13:

- 假设 X^n 为来自标准柯西分布 $\text{Cauchy}(0,1)$ 的 IID 随机样本, 则对所有整数 $n > 0$, 有

$$\bar{X}_n \sim \text{Cauchy}(0,1)$$

- 因此, 即使当 $n \rightarrow \infty$ 时, 样本均值 \bar{X}_n 也不收敛于任何常数。

WLLN 的 IID 假设

- 假设 $X_i = Z, i = 1, 2, \dots, n$ 。即所有的 X_i 服从一个共同的分布 Z 。
- 此时 $\text{corr}(X_i, X_j) = 1$ 。
- $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i = \frac{1}{n} \sum_{i=1}^n Z = Z \not\rightarrow \mu$

例 7.14 [弱大数定律和购买并持有交易策略]：

- 现用一个例子对 WLLN 进行经济解释。
- 金融市场中有一种常用的投资策略称为“购买并持有交易策略 (buy-and-hold trading strategy)”，即投资者某日购买某个资产并长期持有直至售出。
- 问该种交易策略所获得的日平均收益率是多少？

例 7.14 (Cont.):

- 假设 X_i 是资产在第 i 期的收益率，不同时期资产的收益率服从 $\text{IID}(\mu, \sigma^2)$ 。同时假设投资者从第 1 期到第 n 期，总计持有资产共 n 期。则**每一期的平均收益率为样本均值**

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

- 当 $n \rightarrow \infty$ 时，有

$$\bar{X}_n \xrightarrow{p} \mu = E(X_i)$$

- 即当持有期数 n 足够大时，购买并持有交易策略的平均收益率近似等于总体均值 μ 。换言之，总体均值 μ 可视为购买并持有交易策略的长期平均收益率。

例 7.15 [弱大数定律和联合担保贷款]:

- 许多中小企业曾很难申请到银行贷款。
- 为此，不少商业银行推出了一种金融创新产品 —— 联合担保贷款，要求多个中小企业自愿组成一个联保体，若借贷企业破产，则其他企业须帮助偿还贷款。
- 然而，这种针对中小企业的联合担保贷款实际上并不能为银行降低信用风险，因为联合担保企业通常都处于同一区域或同一行业，都面临同一区域或同一行业的共同的系统风险，因此银行无法分散风险。
- 由于同一地区或同一行业的企业之间存在强关联，大数定律失效。

◆ 问题

L_p -收敛和依概率收敛之间的关系是怎样的？



引理 7.4

假设当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{L_p} Z$ 。则当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{p} Z$ 。

证明:

- 由马尔可夫不等式, 若 $\lim_{n \rightarrow \infty} E|Z_n - Z|^p = 0$, 则对所有 $\epsilon > 0$ 有

$$P[|Z_n - Z| > \epsilon] \leq \frac{E|Z_n - Z|^p}{\epsilon^p} \rightarrow 0$$

证毕。

- 注:** 当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{p} Z \not\Rightarrow$ 当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{L_p} Z$

例 7.16:

假设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为 IID $N(\mu, \sigma^2)$ 随机样本。证明 $S_n^2 \xrightarrow{p} \sigma^2$ 。

解：第六章定理 6.6 已证，对正态分布随机样本 \mathbf{X}^n ，以及所有 $n > 1$ ，有

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2$$

- 则 $E(S_n^2) = \sigma^2$ 和 $\text{var}(S_n^2) = 2\sigma^4/(n-1)$ 。
- 因此，当 $n \rightarrow \infty$ 时，有

$$\begin{aligned} E(S_n^2 - \sigma^2)^2 &= \text{var}(S_n^2) \\ &= \frac{2\sigma^4}{n-1} \\ &\rightarrow 0 \end{aligned}$$

- 由引理 7.4 得，当 $n \rightarrow \infty$ 时，有 $S_n^2 \xrightarrow{p} \sigma^2$ 。

例 7.17:

假设一个二元随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 定义如下

Z_n	$\frac{1}{n}$	n
$f_{Z_n}(z_n)$	$1 - \frac{1}{n}$	$\frac{1}{n}$

(1) Z_n 是否依二次方均值收敛于 0?

(2) Z_n 是否依概率收敛于 0?

解:

(1) Z_n 不依二次方均值收敛于 0, 因为

$$\begin{aligned}
 E(Z_n - 0)^2 &= \sum_{Z_n} (z_n - 0)^2 f_{Z_n}(z_n) \\
 &= n^{-2}(1 - n^{-1}) + n^2(n^{-1}) \\
 &> n \rightarrow \infty
 \end{aligned}$$

(2) 给定任意 $\epsilon > 0$, 对所有 $n > N(\epsilon) = [\epsilon^{-1}] + 1$ (故 $n^{-1} < \epsilon$), 有

$$\begin{aligned} P(|Z_n - 0| \leq \epsilon) &= P(Z_n = n^{-1}) \\ &= 1 - n^{-1} \\ &\rightarrow 1, \text{ 当 } n \rightarrow \infty \end{aligned}$$

• 因此, $Z_n \xrightarrow{p} 0$ 。

引理 7.5 [连续性 (Continuity)]

- 假设 $g(\cdot)$ 为连续函数, 且当 $n \rightarrow \infty$ 时, Z_n 依概率收敛于 Z , 则 $g(Z_n)$ 也依概率收敛于 $g(Z)$ 。即若 $g(\cdot)$ 连续, 且当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{p} Z$, 则

$$g(Z_n) \xrightarrow{p} g(Z)$$

或等价地

$$p \lim g(Z_n) = g(p \lim Z_n)$$

证明:

- 根据函数 $g(\cdot)$ 的连续性定义, 对任意 $\epsilon > 0$, 存在常数 $\delta = \delta(\epsilon)$, 使得对任何 $|Z_n - Z| \leq \delta$, 有 $|g(Z_n) - g(Z)| < \epsilon$
- 现定义两个事件

$$A_n(\delta) \equiv \{s \in S: |Z_n(s) - Z(s)| \leq \delta\}$$

$$B_n(\epsilon) \equiv \{s \in S: |g[Z_n(s)] - g[Z(s)]| \leq \epsilon\}$$

- 由 $g(\cdot)$ 的连续性可推出 $A_n(\delta) \subseteq B_n(\epsilon)$, 即 $A_n(\delta)$ 是 $B_n(\epsilon)$ 的一个子集,
- 故有

$$P[A_n(\delta)] \leq P[B_n(\epsilon)]$$

证明 (Cont.):

- 当 $n \rightarrow \infty$ 时

$$P[B_n(\epsilon)^c] \leq P[A_n(\delta)^c] \rightarrow 0$$

其中 $B_n(\epsilon)^c$ 和 $A_n(\delta)^c$ 分别是 $B_n(\epsilon)$ 和 $A_n(\delta)$ 的补集。

- 因为 ϵ 为任意值, δ 亦为任意值。故有

$$\text{当 } n \rightarrow \infty, \quad g(Z_n) \xrightarrow{p} g(Z).$$

证毕。

引理 7.5 [连续性] 的含义

- 上述引理是依概率收敛最重要的性质之一，它表示只要**非线性**函数连续， p lim 运算符就可穿透非线性函数。这类似于微积分中连续函数的极限等于极限的函数值这一性质。当将依概率收敛的随机序列应用于连续函数时，上述引理与微积分中常见的极限性质类似。
- 另一方面， L_p 收敛中使用的期望算子 (expectation operator) $E(\cdot)$ 并不具有该性质，即若 $Z_n \xrightarrow{L_p} Z$ ，不一定有 $g(Z_n) \xrightarrow{L_p} g(Z)$ 。

例 7.18:

令 Z_n 依概率收敛于常数 $c \neq 0$, 证明随机变量 Z_n/c 依概率收敛于 1。

解: 该结果由引理 7.5 以及当 $c \neq 0$ 时函数 $g(z) = z/c$ 为连续函数即可得。

例 7.19:

假设 Z_n 依概率收敛于一常数 c , 并对任意 n 有 $P(Z_n < 0) = 0$ 。证明随机变量 $\sqrt{Z_n}$ 依概率收敛于 \sqrt{c} 。

解: 该结果由引理 7.5 以及平方根函数 $g(z) = \sqrt{z}$ 的连续性即可得。

目 录

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定义 7.8 [几乎处处收敛 (Almost Sure Convergence)]:

- 若对每个给定常数 $\epsilon > 0$, 有

$$P\left(\lim_{n \rightarrow \infty} |Z_n - Z| > \epsilon\right) = 0$$

- 或等价地

$$P\left[S \in S: \lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| \leq \epsilon\right] = 1$$

- 其中 S 为样本空间, 则称随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 几乎处处收敛于随机变量 Z , 记作

$$Z_n \xrightarrow{a.s.} Z, \quad Z_n - Z = o_{a.s.}(1), \quad \text{或} \quad Z_n - Z \xrightarrow{a.s.} 0.$$

几乎处处收敛的含义

- 定义集合

$$\begin{aligned} A(\epsilon) &= \left\{ s \in S : \lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| \leq \epsilon \right\} \\ &= \left\{ s \in S : |Z_n(s) - Z(s)| \leq \epsilon, \text{ 对所有 } n > N(\epsilon, s) \right\} \end{aligned}$$

- $A(\epsilon)$ 是样本空间 S 的一个子集, 其中所有基本事件 s 均满足 $\lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| \leq \epsilon$ 。直觉上, $A(\epsilon)$ 为 S 的一个**收敛子集**。这是因为对每个基本结果 $s \in A(\epsilon)$, 当 $n \rightarrow \infty$ 时样本路径 $\{Z_n(s), n = 1, 2, \dots\}$ 均收敛于 $Z(s)$, 尽管收敛速度可能随 s 不同而有所差异。若几乎处处收敛成立, 则有

$$P\left(\lim_{n \rightarrow \infty} |Z_n - Z| \leq \epsilon\right) = P[A(\epsilon)] = 1$$

- 换言之, 几乎处处收敛要求收敛集 $A(\epsilon)$ 发生的概率为 1。

几乎处处收敛的含义 (Cont.)

- 概率为 1 意味着：对任意给定 $\epsilon > 0$ ，当 n 足够大时，随机变量 Z_n 落在 Z 的 ϵ 邻域内的概率为 1。
- 可能存在非空补集 $A(\epsilon)^c$,

$$A(\epsilon)^c = \left\{ s \in S: \lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| > \epsilon \right\}$$

对其中每个基本结果 s ，当 $n \rightarrow \infty$ 时样本路径 $Z_n(s)$ 不收敛于 $Z(s)$ ，但 $P[A(\epsilon)^c] = 0$ 。

例 7.20:

假设 $S = [0, 1]$ 为样本空间, 其基本结果服从标准均匀分布。定义两个随机变量

$$Z_n(s) = s + s^n \text{ 和 } Z(s) = s$$

证明 $Z_n - Z \xrightarrow{a.s.} 0$ 。

CONTINUE

证明:

- 对每个 $s \in [0, 1)$, 当 $n \rightarrow \infty$ 时, $s^n \rightarrow 0$ 。
- 则对所有 $s \in A(\epsilon) = [0, 1)$, 当 $n \rightarrow \infty$ 时, 有

$$Z_n(s) = s + s^n \rightarrow s = Z(s)$$

- 即对任意 $\epsilon > 0$ 和区间 $[0, 1)$ 中的任意 s , 存在 $N(\epsilon, s) = \lceil \ln \epsilon / \ln s \rceil + 1$ 使得对所有 $n > N(\epsilon, s)$, 有

$$|Z_n(s) - Z(s)| = s^n < \epsilon$$

- 注意 $N(\epsilon, s)$ 依赖于基本结果 s , 这表明 $Z_n(s)$ 的收敛速度依赖于 s 。

CONTINUE

证明 (Cont.):

- 另一方面, 当 $s = 1$, 有

$$Z_n(s) = s + s^n = 1 + 1^n = 2$$

$$Z(s) = s = 1$$

$$|Z_n(s) - Z(s)| = 1 > \epsilon \quad \left(\text{如 } \epsilon = \frac{1}{2} \right)$$

- 存在 $\Lambda = \{1\}$ 满足对所有 n 有 $Z_n(1) = 2$, 使得当 $n \rightarrow \infty$ 时 $Z_n(1) - Z(1) = 1$ 并不趋于零。
- 但 $P(\Lambda) = 0$ 且 $P[A(\epsilon)] = 1$, 因为 s 服从连续分布。
- 故 $Z_n - Z = o_{a.s.}(1)$ 。

几乎处处收敛的数量级 n^α ，其中常数 α 可正可负：

- (1) 若当 $n \rightarrow \infty$ 时，有 $Z_n/n^\alpha \xrightarrow{a.s.} 0$ ，则随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 的**数量级小于** n^α 的概率为 1，记作

$$Z_n = o_{a.s.}(n^\alpha)$$

- (2) 若存在一个常数 $M < \infty$ ，满足 $P(|Z_n/n^\alpha| > M \text{ 当 } n \rightarrow \infty \text{ 时}) = 0$ ，则随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ **数量级不超过** n^α 的概率为 1，记作

$$Z_n = O_{a.s.}(n^\alpha)$$

◆ 问题

依概率收敛与几乎处处收敛之间的关系是怎样的？

- 首先比较二者**表示形式**的不同，**几乎处处收敛**为

$$P\left(\lim_{n \rightarrow \infty} |Z_n - Z| > \epsilon\right) = 0$$

- **依概率收敛**为

$$\lim_{n \rightarrow \infty} P(|Z_n - Z| > \epsilon) = 0$$

- 对**几乎处处收敛**来说, 在收敛集 $A(\epsilon) = \{s \in S : \lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| < \epsilon\}$ 中, 当 $n \rightarrow \infty$ 时, 每个序列 $Z_n(s)$ 均收敛于 $Z(s)$, 且 $A(\epsilon)$ 发生的概率为 1。
- 而对**依概率收敛**来说, 对应 $|Z_n(s) - Z(s)|$ “大偏离” (即大于 ϵ) 的集合 $A_n^c(\epsilon) = \{s : |Z_n(s) - Z(s)| > \epsilon\}$ 的概率可不为 0, 但当 $n \rightarrow \infty$ 时, 其概率收敛于 0。

引理 7.6

若当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{a.s.} Z$, 则 $Z_n \xrightarrow{p} Z$ 。

例 7.21:

- 令样本空间 S 为闭区间 $[0, 1]$, 其基本结果 s 的产生服从标准均匀分布。对所有 $s \in [0, 1]$, 定义随机变量 $Z(s) = s$ 。同时, 对 $n = 1, 2, \dots$, 定义随机变量序列

$$Z_n(s) = \begin{cases} s + s^n, & 0 \leq s \leq 1 - \frac{1}{n} \\ s + 1, & 1 - \frac{1}{n} < s \leq 1 \end{cases}$$

- (1) Z_n 几乎处处收敛于 Z 吗? 请证明;
- (2) Z_n 依概率收敛于 Z 吗? 请证明;
- (3) Z_n 依 L_p 范数收敛于 Z 吗? 请证明。

解:

(1) 考虑集合

$$A^c(\epsilon) = \{s \in S : \lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| > \epsilon\}$$

- 对任意给定 $s \in [0, 1)$, 当 n 充分大即 $n > N(s) = [1/(1 - s)] + 1$ 时, s 将落入区域 $[0, 1 - n^{-1}]$ 。
- 故对任意 $s \in [0, 1)$, 有 $\lim_{n \rightarrow \infty} |Z_n(s) - Z(s)| = \lim_{n \rightarrow \infty} s^n = 0$ 。
因此补集 $A^c(\epsilon)$ 最多只包含一个基本结果, 即 $s = 1$ 。
- 由于基本结果 s 的产生服从连续分布, 故 $P[A^c(\epsilon)] = 0$ 。因此 Z_n 几乎处处收敛于 Z 。

解 (Cont.):

(2)

- 因为几乎处处收敛可推出依概率收敛, 由引理 7.6 可得

$Z_n \xrightarrow{p} Z$ 。或可考察如下集合

$$A_n^c(\epsilon) = \{s \in S : |Z_n(s) - Z(s)| > \epsilon\}$$

- 不失一般性, 假设 $0 < \epsilon < 1$ 。则对每个 n , 区间 $[1 - n^{-1}, 1]$ 包含在 $A_n^c(\epsilon)$ 之中。另外,

- 若 $|Z_n(s) - Z(s)| = s^n > \epsilon$, 则 $s > \epsilon^{\frac{1}{n}}$ 。故有

$$A_n^c(\epsilon) = [\min(1 - n^{-1}, \epsilon^{\frac{1}{n}}), 1]$$

- 当 $n \rightarrow \infty$ 时

$$P[A_n^c(\epsilon)] = 1 - \min(1 - n^{-1}, \epsilon^{\frac{1}{n}}) \rightarrow 0$$

解 (Cont.):

(3)

- 对任意固定 $p > 0$, 当 $n \rightarrow \infty$ 时

$$\begin{aligned} E|Z_n - Z|^p &= \int_0^1 |Z_n(s) - Z(s)|^p ds \\ &= \int_0^{1-n^{-1}} s^{pn} ds + \int_{1-n^{-1}}^1 ds \\ &= \frac{1}{pn+1} \left(1 - \frac{1}{n}\right)^{pn+1} + \frac{1}{n} \\ &\rightarrow 0 \end{aligned}$$

- 因此, 当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{L_p} Z$ 。

例 7.22:

假设 (S, \mathbb{B}, P) 为一概率空间, 其中样本空间 $S = [0, 1]$, \mathbb{B} 是 σ 域, 而 P 是对 S 的连续均匀概率测度。定义如下随机变量的序列:

$$Z_1(s) = 1, 0 \leq s \leq 1$$
$$Z_2(s) = \begin{cases} 1, & 0 \leq s \leq \frac{1}{2} \\ 0, & \frac{1}{2} < s \leq 1 \end{cases}$$

例 7.22 (Cont.):

$$Z_3(s) = \begin{cases} 0, & 0 \leq s \leq \frac{1}{2} \\ 1, & \frac{1}{2} < s \leq 1 \end{cases}$$

$$Z_4(s) = \begin{cases} 1, & 0 \leq s \leq \frac{1}{3} \\ 0, & \frac{1}{3} < s \leq 1 \end{cases}$$

$$Z_5(s) = \begin{cases} 0, & 0 \leq s \leq \frac{1}{3} \\ 1, & \frac{1}{3} < s \leq \frac{2}{3} \\ 0, & \frac{2}{3} < s \leq 1 \end{cases}$$

例 7.22 (Cont.):

$$Z_6(s) = \begin{cases} 0, & 0 \leq s \leq \frac{2}{3} \\ 1, & \frac{2}{3} < s \leq 1 \\ \dots \end{cases}$$

- 证明 $\{Z_n, n = 1, 2, \dots\}$ 依概率收敛于 0, 但不几乎处处收敛于 0。
- 实际上, 对任意给定 $\epsilon > 0$, 收敛集 $A(\epsilon)$ 是一个空集, 即当 $n \rightarrow \infty$ 时, 对任何 $s \in [0, 1]$, 样本路径 $Z_n(s)$ 不收敛于 $Z(s)$ 。

◆ 问题 7.2

几乎处处收敛和 L_p -收敛之间是什么关系?

- 几乎处处收敛 $\not\Rightarrow L_p$ -收敛,
- L_p -收敛 $\not\Rightarrow$ 几乎处处收敛。

例 7.24:

- 令样本空间 S 为闭区间 $[0, 1]$, 其基本结果 s 的发生服从标准均匀分布。对 $n = 1, 2, \dots$, 定义一个随机变量序列

$$Z_n(s) = \begin{cases} 0, & s \in [0, 1 - n^{-2}] \\ e^n, & s \in (1 - n^{-2}, 1] \end{cases}$$

- 回答如下问题并证明: 当 $n \rightarrow \infty$ 时,

$$(1) Z_n \xrightarrow{q.m} 0? \quad (2) Z_n \xrightarrow{p} 0? \quad (3) Z_n \xrightarrow{a.s.} 0?$$

解:

- (1) 否; (2) 是; (3) 是。

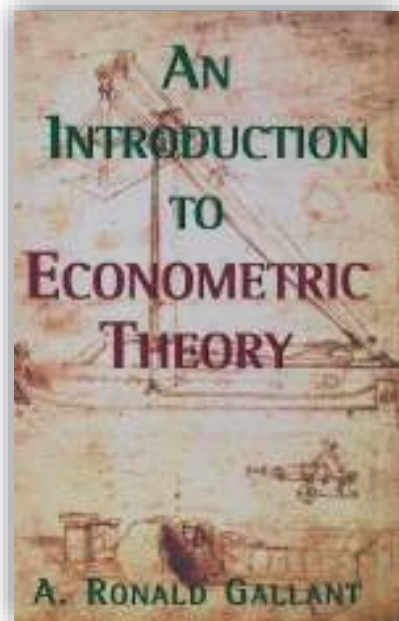
引理 7.7 [连续性 (Continuity)]

假设 $g(\cdot)$ 为连续函数, 且当 $n \rightarrow \infty$ 时, Z_n 几乎处处收敛于 Z 。
则当 $n \rightarrow \infty$ 时, $g(Z_n)$ 几乎处处收敛于 $g(Z)$ 。

定理 7.2 [柯尔莫哥洛夫强大数定律 (Kolmogorov SLLN)]:

- 设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 是从 $E(X_i) = \mu$ 且 $E|X_i| < \infty$ 的总体分布产生的 IID 随机样本。定义样本均值 $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ 。
- 则当 $n \rightarrow \infty$ 时,

$$\bar{X}_n \xrightarrow{a.s.} \mu$$



证明： 参见 Gallant, A. R. (1997), An Introduction to Econometric Theory. Princeton: Princeton University Press. pp.132-135.

定理 7.3 [一致强大数定律 (USLLN)]

• 假设:

(1) $\mathbf{X}^n = (X_1, \dots, X_n)$ 为 IID 随机样本;

(2) 函数 $g(x, \theta)$ 在 $\Omega \times \Theta$ 上**连续**, 其中 Ω 是 X_i 的支撑, Θ 是 \mathbb{R}^d 上的**紧集** (compact set), d 为有限固定整数;

(3) $E[\sup_{\theta \in \Theta} |g(X_i, \theta)|] < \infty$, 其中期望 $E(\cdot)$ 在 X_i 的总体分布上取期望值。

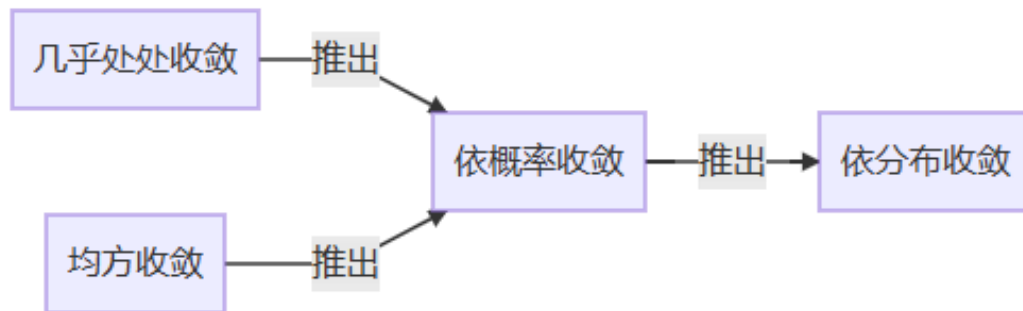
• 则当 $n \rightarrow \infty$ 时

$$\sup_{\theta \in \Theta} \left| n^{-1} \sum_{i=1}^n g(X_i, \theta) - E[g(X_i, \theta)] \right| \xrightarrow{a.s.} 0$$

• 此外, $E[g(X_i, \theta)]$ 是参数空间 Θ 上关于 θ 的连续函数。

收敛概念总结

- $Z_n - Z \xrightarrow{a.s.} 0$: $P\left[\lim_{n \rightarrow \infty} |Z_n - Z| = 0\right] = 1$.
- $Z_n \xrightarrow{L_p} Z$: $\lim_{n \rightarrow \infty} E|Z_n - Z|^p = 0$.
 ✓ $Z_n - Z \xrightarrow{q.m.} 0$: $E(Z_n - Z)^2 \rightarrow 0$ as $n \rightarrow \infty$.
- $Z_n - Z \xrightarrow{p} 0$ 或 $Z_n - Z = o_P(1)$: 对每个常数 $\epsilon > 0$,
 $P[|Z_n - Z| > \epsilon] \rightarrow 0$ as $n \rightarrow \infty$ 或
 $P[|Z_n - Z| \leq \epsilon] \rightarrow 1$ as $n \rightarrow \infty$.
- $Z_n \xrightarrow{d} Z$: 对任意连续点 z $\lim_{n \rightarrow \infty} F_n(z) = F(z)$ or $F_n(z) \rightarrow F(z)$ as $n \rightarrow \infty$.



第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定义 7.9[依分布收敛 (Convergence in Distribution)]:

- 令 $\{Z_n, n = 1, 2, \dots\}$ 为随机变量序列, 其对应的 CDF 序列为 $\{F_n(z), n = 1, 2, \dots\}$,
- 令随机变量 Z 的 CDF 为 $F(z)$ 。
- 若在 $F(z)$ 的每个连续点 $z \in (-\infty, \infty)$ 处有 CDF $F_n(z)$ 收敛于 $F(z)$, 即在每个 $F(z)$ 连续的 z 点处都有

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F(z)$$

- 则称当 $n \rightarrow \infty$ 时 Z_n 依分布收敛于 Z , 其中 $F(z)$ 为随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 的**极限 (或渐近) 分布 (limiting or asymptotic distribution)**。

依分布收敛是关于累积分布函数 CDF 的收敛概念

- 尽管称随机变量序列 $\{Z_n, n = 1, 2, \dots\}$ 依分布收敛于随机变量 Z , 其实是指累积分布函数 CDF 序列 $\{F_n(z), n = 1, 2, \dots\}$ 收敛于 CDF $F(\cdot)$ 。
- 依分布收敛:
 - ✓ $F_n(z)$ 与 $F(\cdot)$ 的接近程度
- 依 L_p 收敛、依概率收敛以及几乎处处收敛:
 - ✓ Z_n 与 Z 的接近程度

极限分布 $F(\cdot)$ 可能无法由 $F_n(\cdot)$ 求极限获得

- 例如, 若 $Z_n \sim N\left(0, \frac{1}{n}\right)$, 则其分布函数

$$\begin{aligned} F_n(z) &= \int_{-\infty}^z \frac{1}{\sqrt{1/n}\sqrt{2\pi}} e^{-nu^2/2} du \\ &= \int_{-\infty}^{\sqrt{n}z} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} dv = \Phi(\sqrt{n}z) \end{aligned}$$

其中 $\Phi(\cdot)$ 是 $N(0, 1)$ 的 CDF。

- 显然, 有

$$\lim_{n \rightarrow \infty} F_n(z) = \begin{cases} 0, & z < 0 \\ \frac{1}{2}, & z = 0 \\ 1, & z > 0 \end{cases}$$

极限分布 $F(\cdot)$ 可能无法由 $F_n(\cdot)$ 求极限获得 (Cont.)

- 现定义函数

$$F(z) = \begin{cases} 0, & z < 0 \\ 1, & z \geq 0 \end{cases}$$

则 $F(z)$ 为 CDF 且在 $F(z)$ 的每个连续点处有 $\lim_{n \rightarrow \infty} F_n(z) = F(z)$ (函数 $F(z)$ 只在 $z = 0$ 处不连续)。

- 因此, $F(\cdot)$ 是 Z_n 的极限分布。但 $F(\cdot)$ 无法通过对 $F_n(\cdot)$ 求极限得到, 因为在 $z = 0$ 点处 $\lim_{n \rightarrow \infty} F_n(z) = F_n(0) \neq F(0)$ 。

例 7.25:

设 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为 IID $U[0, \theta]$ 随机样本, 其中 θ 是未知参数。

定义 $Z_n = \max_{1 \leq i \leq n} (X_i)$ 为 θ 的估计量。推导 $n(\theta - Z_n)$ 的极限分布。

解:

- 任意给定 $u \geq 0$, 当 $n \rightarrow \infty$ 时, 有

$$\begin{aligned}
 P[n(\theta - Z_n) > u] &= P\left(Z_n < \theta - \frac{u}{n}\right) \\
 &= P\left(X_1 < \theta - \frac{u}{n}, \dots, X_n < \theta - \frac{u}{n}\right) \\
 &= \prod_{i=1}^n P\left(X_i < \theta - \frac{u}{n}\right) \\
 &= \left(1 - \frac{u}{n\theta}\right)^n \\
 &\rightarrow e^{-\frac{u}{\theta}} \quad \left[\left(1 - \frac{a}{n}\right)^n \rightarrow e^{-a}\right]
 \end{aligned}$$

- 则对任意 $u \geq 0$, 当 $n \rightarrow \infty$ 时

$$F_n(u) = 1 - P[n(\theta - Z_n) > u] \rightarrow 1 - e^{-u/\theta}$$

- 这说明 $n(\theta - Z_n)$ 依分布收敛于参数为 θ 的指数分布 (exponential distribution) $EXP(\theta)$ 。

◆ 问题

依分布收敛和依概率收敛之间是什么关系？

- 依 L_p 收敛、依概率收敛以及几乎处处收敛：
 - ✓ Z_n 与 Z 的接近程度
- 依分布收敛：
 - ✓ $F_n(z)$ 与 $F(\cdot)$ 的接近程度
 - ✓ 两个随机变量可以不靠近，但他们的分布可以是一样的。

引理 7.8

令 Z_n 是 CDF 为 $F_n(\cdot)$ 的随机变量, Z 是 CDF 为 $F(\cdot)$ 的连续随机变量。若当 $n \rightarrow \infty$ 时 $Z_n \xrightarrow{d} Z$, 则 $Z_n = O_p(1)$ 。

证明:

- 对任意给定常数 $\epsilon > 0$, 令 $M = M(\epsilon)$ 为满足 $P(|Z| > M) < \epsilon$ 的常数。
- $F_n(z)$ 是 Z_n 的 CDF。
- 给定 $Z_n \xrightarrow{d} Z$ 且 Z 的 CDF $F(z)$ 处处连续, 则对任意 $z \in (-\infty, \infty)$ 和足够大的 n , 有 $|F_n(z) - F(z)| \leq \epsilon$ 。

证明 (Cont.):

- 这说明对所有充分大的 n , 有

$$P(Z_n > M) - P(Z > M) \leq \epsilon$$

$$P(Z_n \leq -M) - P(Z \leq -M) \leq \epsilon$$

- 则

$$\begin{aligned} P(Z_n > M) + P(Z_n < -M) &\leq P(Z_n > M) + P(Z_n \leq -M) \\ &< P(Z > M) + P(Z \leq -M) + 2\epsilon \\ &= P(Z > M) + P(Z < -M) + 2\epsilon \end{aligned}$$

- 这里给定 Z 为连续随机变量, 故有 $P(Z = -M) = 0$ 。因此

$$P(|Z_n| > M) < P(|Z| > M) + 2\epsilon < 3\epsilon \equiv \delta$$

- 因为 ϵ 是任意值, δ 也是任意值, 故 $Z_n = O_p(1)$ 。

证毕。

例 7.26:

- 在例 7.25 中, 统计量 $Z_n = \max_{1 \leq i \leq n} (X_i)$, 其中 $\mathbf{X}^n = (X_1, \dots, X_n)$ 是来自均匀分布 $U[0, \theta]$ 的 IID 随机样本, 例 7.25 已证明当 $n \rightarrow \infty$ 时, $n(\theta - Z_n) \xrightarrow{d} EXP(\theta)$ 。
- 因此, $n(\theta - Z_n) = O_p(1)$, 或 $Z_n - \theta = O_p(n^{-1})$ 。这表明 Z_n 以速度 n^{-1} 依概率逼近收敛于 θ 。
- 直觉上, 随机变量的 n 个观测值 $\{X_i\}_{i=1}^n$ 或多或少地均匀分布于整个支撑区间 $[0, \theta]$ 。
- 因此, $\{X_i\}_{i=1}^n$ 的最大值将以速度 n^{-1} 逼近均匀分布的上界 θ 。

引理 7.9

假设当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{p} Z$ 。则当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{d} Z$ 。

- **证明:** 留作练习题。

引理 7.10 [渐近等价性 (Asymptotic Equivalence)]

假设当 $n \rightarrow \infty$ 时, $Y_n - Z_n \xrightarrow{p} 0$ 且 $Z_n \xrightarrow{d} Z$ 。则当 $n \rightarrow \infty$, $Y_n \xrightarrow{d} Z$ 。

引理 7.10 的应用

- 若 $X_i \sim \text{IID } N(\mu_0, \sigma^2)$, 可以得到

$$y_n = T(x^n) = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{S_n} \sim t_{n-1}$$

- 若正态分布的假设不再成立, 上面结论不再成立。

- 为此, 我们构建 $z_n = \frac{\sqrt{n}(\bar{X}_n - \mu_0)}{\sigma}$, 根据中心极限定理, 可以推

得当 $n \rightarrow \infty$ 时, $z_n \xrightarrow{d} N(0,1)$ 。

- 同时, 可以得到 $y_n - z_n = \frac{\sqrt{n}(\bar{x}_n - \mu_0)}{\sigma} \left[\frac{\sigma}{S_n} - 1 \right] = O_p(1) o_p(1) \xrightarrow{p} 0$ 。

- 利用引理 7.10 我们可以得到 $y_n \xrightarrow{d} N(0,1)$ 。

定理 7.5 [连续映射定理 (Continuous Mapping Theorem)]

- 假设当 $n \rightarrow \infty$ 时, k 维随机向量序列 $Z_n \xrightarrow{d} Z$ 且 $g: \mathbb{R}^k \rightarrow \mathbb{R}^l$ 为 l 维连续向量值函数, 其中 k, l 是有限且固定正整数。
- 则当 $n \rightarrow \infty$ 时, 有 $g(Z_n) \xrightarrow{d} g(Z)$ 。

目 录

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

定理 7.6 [独立同分布随机样本的林德伯格-列维中心极限定理 (Lindeberg-Levy's Central Limit Theorem, CLT)]:

- 令 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为来自均值为 μ , 方差为 σ^2 ($0 < \sigma^2 < \infty$) 的某个总体分布的 IID 随机样本。
- 定义样本均值 $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$ 。
- 当 $n \rightarrow \infty$ 时, 则标准化样本均值

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\text{var}(\bar{X}_n)}} \\ &= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \\ &= \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \\ &\stackrel{d}{\rightarrow} N(0,1) \end{aligned}$$

定理 7.6 (Cont.)

- 标准正态随机变量 $Z \sim N(0, 1)$ 的 CDF 常记作

$$\Phi(z) = \int_{-\infty}^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

- CLT 表明, 当 $n \rightarrow \infty$ 时, $Z_n \xrightarrow{d} Z$, 即对所有 $z \in (-\infty, \infty)$, $F_n(z) \rightarrow \Phi(z)$, 其中 $F_n(z)$ 是 Z_n 的 CDF。

证明:

- 定义标准化随机变量

$$Y_i = \frac{X_i - \mu}{\sigma}, i = 1, \dots, n$$

- 其特征函数为 $\varphi_Y(t) = E(e^{itY_i})$, 其中 $\mathbf{i} = \sqrt{-1}$.
- 则 Y_i 均值为 0, 方差为 1。有

$$\begin{aligned}\varphi_Y'(0) &= 0 \\ \varphi_Y''(0) &= \mathbf{i}^2 \sigma_Y^2 = -1\end{aligned}$$

证明 (Cont.):

- 标准化样本均值为

$$\begin{aligned} Z_n &= \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \\ &= \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \\ &= \sqrt{n} \left(n^{-1} \sum_{i=1}^n \frac{X_i - \mu}{\sigma} \right) \\ &= \sqrt{n} \bar{Y}_n \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i \end{aligned}$$

- 由于 X_i 的矩生成函数 MGF 不一定存在, 因此将采用特征函数法, 证明当 $n \rightarrow \infty$ 时, $\varphi_n(t) \rightarrow e^{-\frac{t^2}{2}}$, 其中 $\varphi_n(t) = E(e^{itZ_n})$ 是 Z_n 的特征函数, 而 $e^{-\frac{t^2}{2}}$ 是 $N(0, 1)$ 的特征函数。

证明 (Cont.):

- 给定 IID 假设, 有

$$\begin{aligned}\varphi_n(t) &= E(e^{itZ_n}) \\ &= E\left(e^{it\sqrt{n}\bar{Y}_n}\right) \\ &= E\left(e^{\frac{it}{\sqrt{n}}\sum_{i=1}^n Y_i}\right) \\ &= E\left(e^{\frac{it}{\sqrt{n}}Y_1} e^{\frac{it}{\sqrt{n}}Y_2} \dots e^{\frac{it}{\sqrt{n}}Y_n}\right) \\ &= E\left(e^{\frac{it}{\sqrt{n}}Y_1}\right) E\left(e^{\frac{it}{\sqrt{n}}Y_2}\right) \dots E\left(e^{\frac{it}{\sqrt{n}}Y_n}\right) \\ &= \left[E\left(e^{\frac{it}{\sqrt{n}}Y_1}\right)\right]^n \\ &= \left[\varphi_Y\left(\frac{t}{\sqrt{n}}\right)\right]^n\end{aligned}$$

- 其中 $\varphi_Y(t) = E(e^{itY_i})$ 。

证明 (Cont.):

- 现在, 有

$$\begin{aligned}\ln \left\{ [\varphi_Y(t/\sqrt{n})]^n \right\} &= n \ln [\varphi_Y(t/\sqrt{n})] \\ &= \frac{\ln [\varphi_Y(t/\sqrt{n})]}{1/n}\end{aligned}$$

- 给定 $\varphi_Y(0) = 1$ 和 $1/n \rightarrow 0$, $\ln[\varphi_Y(t/\sqrt{n})] \rightarrow 0$, 对任意给定 $t \in (-\infty, \infty)$, 由洛必达法则 (L'Hospital's rule) 可得:

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{\ln [\varphi_Y(t/\sqrt{n})]}{1/n} &= \lim_{n \rightarrow \infty} \frac{\frac{\varphi_Y'(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})} \left(-\frac{t}{2n\sqrt{n}} \right)}{-1/n^2} \\ &= \frac{t}{2} \lim_{n \rightarrow \infty} \frac{\frac{\varphi_Y'(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})}}{1/\sqrt{n}}\end{aligned}$$

证明 (Cont.):

- 其中

$$\frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})} \rightarrow 0$$

- 给定 $\varphi'_Y(0) = 0$ 和 $1/\sqrt{n} \rightarrow 0$, 再次应用洛必达法则得

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\varphi'_Y(t/\sqrt{n})}{\varphi_Y(t/\sqrt{n})} &= \lim_{n \rightarrow \infty} \frac{\frac{\varphi''_Y(t/\sqrt{n})\varphi_Y(t/\sqrt{n}) - [\varphi'_Y(t/\sqrt{n})]^2}{[\varphi_Y(t/\sqrt{n})]^2} \left(-\frac{t}{2n\sqrt{n}}\right)}{-\frac{1}{2n\sqrt{n}}} \\ &= t \lim_{n \rightarrow \infty} \frac{\varphi''_Y(t/\sqrt{n})\varphi_Y(t/\sqrt{n}) - [\varphi'_Y(t/\sqrt{n})]^2}{[\varphi_Y(t/\sqrt{n})]^2} \\ &= -t \end{aligned}$$

- 其中用到 $\varphi_Y(0) = 1$, $\varphi'_Y(0) = 0$, $\varphi''_Y(0) = -1$.

证明 (Cont.):

- 则有

$$\begin{aligned}\lim_{n \rightarrow \infty} \ln \varphi_n(t) &= \lim_{n \rightarrow \infty} \frac{\ln[\varphi_Y(t/\sqrt{n})]}{1/n} \\ &= -\frac{1}{2}t^2\end{aligned}$$

- 因为连续函数的极限 (此处为指数函数) 等于极限的函数, 有

$$\lim_{n \rightarrow \infty} \varphi_n(t) = e^{-\frac{t^2}{2}}$$

- 因此, 当 $n \rightarrow \infty$ 时 $Z_n \xrightarrow{d} N(0, 1)$ 。

证毕。

另一种启发式证明:

- Z_n 的特征函数为

$$\begin{aligned}
 \varphi_n(t) &= E\left(e^{it\sqrt{n}\bar{Y}_n}\right) \\
 &= \left[E\left(e^{itY_1/\sqrt{n}}\right)\right]^n \\
 &= \left[\varphi_Y(t/\sqrt{n})\right]^n \\
 &= \left[\varphi_Y(0) + \varphi_Y'(0)\frac{t}{\sqrt{n}} + \frac{1}{2}\varphi_Y''(0)\left(\frac{t}{\sqrt{n}}\right)^2 + r\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
 &= \left[1 - \frac{t^2}{2n} + o(n^{-1})\right]^n \\
 &\rightarrow e^{-t^2/2}
 \end{aligned}$$

- 其中 $r(t/\sqrt{n})$ 代表余项, 且应用了极限公式: 当 $n \rightarrow \infty$ 时, $\left(1 + \frac{a}{n}\right)^n \rightarrow e^a$ 。

另一种启发式证明 (Cont.):

- 这个公式在第四章推导**小数定律**时也发挥了重要作用, 即一些服从伯努利分布 (p) 的独立随机变量, 它们的和服从二项分布 $B(n, p)$, 而当 $np \rightarrow \lambda$ 且 $n \rightarrow \infty$ 时, 和的分布便趋于泊松分布 $\text{Poisson}(\lambda)$ 。

例 7.29 [二项分布 $B(n, p)$ 的正态近似 (Normal Approximation)]:

- 服从二项分布 $B(n, p)$ 的随机变量 Z_n , 可表示为 $Z_n = \sum_{i=1}^n X_i$, 其中 $\mathbf{X}^n = (X_1, \dots, X_n)$ 为来自伯努利分布 $\text{Bernoulli}(p)$ 的 IID 随机样本, 且 $p = P(X_i = 1)$.

- 由 CLT, 有 $n \rightarrow \infty$ 时, 标准化随机变量

$$\frac{Z_n - E(Z_n)}{\sqrt{\text{var}(Z_n)}} = \frac{Z_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

- 虽然该结果仅当 $n \rightarrow \infty$ 时适用, 但在实际应用中, 即使样本量 n 不太大, 有时仍使用正态分布近似二项分布。

例 7.30 [χ_n^2 的正态近似]:

- 设 X^n 为 IID $N(0, 1)$ 随机样本。令 $Y_i = X_i^2, i = 1, \dots, n$ 。
- 则当 $n \rightarrow \infty$ 时, 标准化随机变量

$$\frac{\sum_{i=1}^n Y_i - n\mu_Y}{\sqrt{n\sigma_Y^2}} = \frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}}$$
$$\xrightarrow{d} N(0, 1)$$

解:

- 令 $\bar{Y}_n = n^{-1} \sum_{i=1}^n Y_i$ 。因为 $E(Y_i) = 1$ 和 $\text{var}(Y_i) = 2$ ，根据 CLT，当 $n \rightarrow \infty$ 时，标准化样本均值

$$\frac{\bar{Y}_n - \mu_Y}{\sigma_Y/\sqrt{n}} = \frac{\bar{Y}_n - 1}{\sqrt{2}/\sqrt{n}} \xrightarrow{d} N(0,1)$$

- 其中

$$\begin{aligned} \frac{\bar{Y}_n - 1}{\sqrt{2}/\sqrt{n}} &= \frac{\sqrt{n}(\bar{Y}_n - 1)}{\sqrt{2}} \\ &= \frac{\frac{1}{\sqrt{n}} \sum_{i=1}^n (Y_i - 1)}{\sqrt{2}} \\ &= \frac{\sum_{i=1}^n Y_i - n}{\sqrt{2n}} \\ &= \frac{\sum_{i=1}^n X_i^2 - n}{\sqrt{2n}} \end{aligned}$$

解 (Cont.):

- 由引理 6.1, $\sum_{i=1}^n X_i^2 \sim \chi_n^2$ 。因此, 当自由度 n 足够大时, 可用 $N(n, 2n)$ 分布近似 χ_n^2 分布。
- 例如, 当 \mathbf{X}^n 为 IID $N(\mu, \sigma^2)$ 随机样本时, 对所有 $n > 1$ 有

$$\frac{(n-1)S_n^2}{\sigma^2} \sim \chi_{n-1}^2。$$

- 则当 $n \rightarrow \infty$ 时,

$$\left[\frac{(n-1)S_n^2}{\sigma^2} - (n-1) \right] / \sqrt{2(n-1)} \xrightarrow{d} N(0,1)$$

◆ 问题 7.3

CLT 中有限方差假设 $\text{var}(X_i) = \sigma^2 < \infty$ 的重要性如何呢?

例 7.31 [独立柯西随机变量之和]:

- 假设 X^n 为来自标准柯西分布 $\text{Cauchy}(0, 1)$ 的 IID 随机样本。
- 证明对所有 $n \geq 1$, $\bar{X}_n \sim \text{Cauchy}(0, 1)$ 。

解:

- 现用特征函数证明该结论。由 IID 性质和第四章给出的 Cauchy(0, 1) 随机变量的特征函数 $\varphi(t) = e^{-|t|}$, 有

$$\begin{aligned}\varphi_{\bar{X}_n}(t) &= E(e^{it\bar{X}_n}) \\ &= \left[\varphi\left(\frac{t}{n}\right) \right]^n \\ &= e^{-|t|} \\ &= \varphi(t)\end{aligned}$$

- 因此, 对所有 $n \geq 1$ 有 $\bar{X}_n \sim \text{Cauchy}(0, 1)$ 。
- 当 $n \rightarrow \infty$ 时, 一个 IID Cauchy(0, 1) 随机样本的样本均值不收敛于 $N(0, 1)$ 。

定理 7.7 [独立非同分布随机样本的李雅普诺夫中心极限定理 (Liapounov's (1901) CLT for Independent Random Variables)]

设随机变量 X_1, \dots, X_n 联合独立, 且对 $i = 1, \dots, n$ 有 $E|X_i - \mu_i|^3 < \infty$, 其中 $E(X_i) = \mu_i$ 。同时, 假设

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n E|X_i - \mu_i|^3}{\left(\sum_{i=1}^n \sigma_i^2\right)^{3/2}} = 0$$

则当 $n \rightarrow \infty$ 时, 标准化随机变量

$$Z_n = \frac{\sum_{i=1}^n X_i - \sum_{i=1}^n \mu_i}{\left(\sum_{i=1}^n \sigma_i^2\right)^{1/2}} \xrightarrow{d} N(0,1)$$

- 当随机变量 X_1, \dots, X_n 之间存在一定程度的相依 (dependence) 时, CLT 仍然成立。
- 生成时间序列数据的随机样本一般会出现这种相依性。尽管这种相依性不能太强 (考虑在 $X_1 = X_2 = \dots = X_n$ 的极端情况下会出现什么问题), 但仍可一定程度上放松独立性假设。

定理 7.8 [斯勒茨基 (Slutsky) 定理]:

- 假设当 $n \rightarrow \infty$ 时, $X_n \xrightarrow{d} X$ 和 $C_n \xrightarrow{p} c$, 其中 c 为常数。
- 则当 $n \rightarrow \infty$ 时, 有
 - (1) $X_n + C_n \xrightarrow{d} X + c$;
 - (2) $X_n - C_n \xrightarrow{d} X - c$;
 - (3) $X_n C_n \xrightarrow{d} cX$;
 - (4) 若 $c \neq 0$, $\frac{X_n}{C_n} \xrightarrow{d} \frac{X}{c}$ 。

证明: 留作练习题 (应用依概率渐近等价引理, 即引理 7.10)。

例 7.32

- 假设标准化样本均值

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} N(0,1)$$

且当 $n \rightarrow \infty$ 时, $S_n^2 \xrightarrow{p} \sigma^2$ 。

- 故由平方根函数的连续性有 $S_n \xrightarrow{p} \sigma$ 。
- 则由斯勒茨基定理可得

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \frac{\sigma}{S_n}$$
$$\xrightarrow{d} N(0,1)$$

例 7.33

假设当 $n \rightarrow \infty$ 时, $X_n \xrightarrow{d} X$ 和 $Y_n \xrightarrow{d} Y$ 。以下结果是否成立? 给出推理过程。

(1) $X_n \pm Y_n \xrightarrow{d} X \pm Y$, 当 $n \rightarrow \infty$ 时;

(2) $X_n Y_n \xrightarrow{d} XY$, 当 $n \rightarrow \infty$ 时。

解:

- 一般而言, 答案是**否定的**, 因为它们并未考虑 X_n 和 Y_n 之间的相依性。换言之, 依边际分布收敛不一定能推出依联合分布收敛。这不同于其他收敛性概念如依二次方均值收敛、依概率收敛以及几乎处处收敛。

引理 7.11 [德尔塔 (Delta) 方法]:

假设当 $n \rightarrow \infty$ 时, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$, 同时 $g(\cdot)$ 为连续可导函数且 $g'(\mu) \neq 0$ 。则当 $n \rightarrow \infty$ 时

$$\sqrt{n}[g(\bar{X}_n) - g(\mu)] \xrightarrow{d} N(0, \sigma^2 [g'(\mu)]^2)$$

和

$$\frac{\sqrt{n}[g(\bar{X}_n) - g(\mu)]}{\sigma g'(\mu)} \xrightarrow{d} N(0,1)$$

证明:

- 首先, 由引理 7.8, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$ 可推出 $\sqrt{n}(\bar{X}_n - \mu)/\sigma = O_p(1)$ 。因此 $\bar{X}_n - \mu = O_p(n^{-1/2}) = o_p(1)$ 。
- 其次, 根据中值定理, 有

$$Y_n = g(\bar{X}_n) = g(\mu) + g'(\bar{\mu}_n)(\bar{X}_n - \mu)$$

其中存在 $\lambda \in [0, 1]$, $\bar{\mu}_n = \lambda\mu + (1 - \lambda)\bar{X}_n$, 注意

$$|\bar{\mu}_n - \mu| = |(1 - \lambda)(\bar{X}_n - \mu)| \leq |\bar{X}_n - \mu| = o_p(1)$$

证明 (Cont.):

- 根据斯勒茨基定理, 有

$$\sqrt{n} \left[\frac{g(\bar{X}_n) - g(\mu)}{\sigma} \right] = g'(\bar{\mu}_n) \sqrt{n} \frac{\bar{X}_n - \mu}{\sigma} \\ \xrightarrow{d} N[0, g'(\mu)^2]$$

- 其中, 给定 $\bar{\mu}_n \xrightarrow{p} \mu$ 和一阶导数 $g'(\cdot)$ 的连续性, 由引理 7.5 可得 $g'(\bar{\mu}_n) \xrightarrow{p} g'(\mu)$ 。
- 又据斯勒茨基定理, 有

$$\frac{\sqrt{n}[g(\bar{X}_n) - g(\mu)]}{\sigma g'(\bar{X}_n)} \xrightarrow{d} N(0,1)$$

证毕。

目 录

第一节 极限和数量级

第二节 收敛概念的必要性

第三节 依二次方均值收敛和 L_p 收敛

第四节 依概率收敛

第五节 几乎处处收敛

第六节 依分布收敛

第七节 中心极限定理

第八节 小结

- **极限、连续性和数量级**等数学概念
- 渐近理论的基础概念与分析工具
 - ✓ **四个收敛概念：**
 - 依二次方均值收敛、依概率收敛、几乎处处收敛：刻画 Z_n 与 Z 之间的接近程度
 - 依分布收敛：刻画 $F_n(z)$ 与 $F(\cdot)$ 之间的接近程度
- **大数定理、中心极限定理、Slutsky 定理和 Delta 方法**



中国科学院数学与系统科学研究院

Academy of Mathematics and Systems Science

Chinese Academy of Sciences

Thank You !