计量经济学

第3讲：统计复习

授课人：刘　岩

中山大学商学院

2025年3月3日

本讲内容

1. 统计概要
2. 常见估计方法
3. 统计推断基础

本节内容

1 统计概要

2 常见估计方法

3 统计推断基础

样本与总体：1-维情形

- 假设某个经济、金融变量为随机变量，分布为 F，有 n 个观测值 (observation) $X_1, \ldots, X_n$
- 称这组观测值为数据样本 (data sample)，简称样本
- 称 F 为总体分布 (population distribution)，简称总体
  - 例如 GDP 增速，每个季度的观测值就是样本；季度 GDP 增速的具体取值，如 2020Q4 的 6.8/7.2，称为样本的实现值 (realized value/realization)
- 统计学的出发点：数据样本总是从总体分布中抽样 (sampling) 得到
  - 从总体中的抽样过程，又称为数据生成过程 (data generating process, DGP)

样本与总体：多维情形

- 多个经济、金融变量，构成随机向量，联合分布为 F，有 n 个观测值 (observation) $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$
- 类似的，称这组向量观测值为<u>样本</u>，称 F 为<u>总体</u>
- 多个变量之间的相互关联，体现在总体分布 F 上
  - 每个季度的 GDP、消费、投资、政府支出、进口、出口满足下列关系

$$Y_t = C_t + I_t + G_t + EX_t - IM_t,$$

  则 $\boldsymbol{X}_t = [Y_t, C_t, I_t, G_t, EX_t, IM_t]^{\mathsf{T}}$ 为一个样本
- 与 1-维情形相同，样本从总体分布中抽样得到

样本的两重涵义

做为随机变量的样本

- 统计模型假设一组随机变量满足一个总体联合分布，按照这个分布，DGP 产生了一系列样本（观测值）$\mathcal{X} = \{\boldsymbol{X}_i\}_{i=1}^n$
- 这一系列样本仍然视作随机变量：符合总体分布 F

作为观测取值的样本

- 在具体的应用中，数据 (data) 指样本 $\{\boldsymbol{X}_i\}_{i=1}^n$ 的具体实现值(realized value, realization)，或样本取值
- 有时为了明确区分样本随机变量和样本取值，将后者写作小写字母 $\{\boldsymbol{x}_i\}_{i=1}^n$

统计的基本问题

- 假设有一个总体 F，但部分（或全部）特征未知
- 研究者观察到一组样本 $\mathcal{X}=\{\mathbf{X}_i\}_{1\le i\le n}$，并且知道这组样本来自 F，亦即这组样本对应的总体是 F

问题 1 如何从观察到的样本 $\mathcal{X}$ 估计 (estimate) F 的部分（或全部）特征？

问题 2 如果研究猜想 F 满足某种特征，如何从 $\mathcal{X}$ 来推断 (infer) 这个猜想的对错？

统计估计与推断的基本概念

- 假设样本的 $\{X_i\}_{i=1}^N$ 所满足的统计模型为给定的总体分布

$$M_\theta = \mathbb{P}\left(\{X_i\}_{i=1}^N \middle| \theta\right),$$

但总体分布的具体性质未知，表示为参数 θ 真实值 θ_0 未知
- 基本任务：从样本 $\{X_i\}_{i=1}^N$ 推测未知的真实值 θ_0
- 估计：从样本 $\{X_i\}_{i=1}^N$ 获得 θ_0 的估计值 $\hat{\theta}_N$
- 推断：由样本计算得到的 $\hat{\theta}_N$ 判断 θ 是否为特定取值 θ_0
- 误差：由于 $\hat{\theta}_N$ 继承了样本的随机性，上述推断可能存在误差 $\Longrightarrow$ 统计推断需要控制误差

本节内容

1 统计概要

2 常见估计方法
- 矩估计
- 极大似然估计
- 估计值的性质

3 统计推断基础

矩估计基础

- 矩估计 (moment estimation) 的想法：由样本 $\mathcal{X} = \{\boldsymbol{X}_i\}$ 计算样本矩 (sample moment) $\hat{m}(\mathcal{X})$，同时选择总体分布的参数 $\boldsymbol{\theta}$ 使得总体矩 (population moment) 与样本矩“最为”接近
  - 参数的矩估计值 $\hat{\boldsymbol{\theta}}$ 是样本的函数：$\hat{\boldsymbol{\theta}}(\mathcal{X})$，常简记为 $\hat{\boldsymbol{\theta}}_N$
- 示例：假设 $\{X_i\}_{i=1}^N \overset{\text{iid}}{\sim} U([0,a])$，$a$ 未知，则可以利用 $\{X_i\}$ 的样本均值来估计 a
  - X_i 的总体均值为 $a/2$，样本均值为 $\hat{\mu}_N = \frac{1}{N}\sum_i X_i$
  - a 的一个矩估计为 $\hat{a} = 2\hat{\mu}_N = \frac{2}{N}\sum_i X_i$
  - 也可以使用高阶矩进行估计
- 总体矩和样本矩之间的“距离”可以通过多种方式进行度量，如误差平方和

极大似然估计基础

- 假设样本 $\mathcal{X} = \{X_i\}_{i=1}^N$ 具有联合分布 $\mathbb{P}(\mathcal{X}|\boldsymbol{\theta})$
- 将给定样本取值时，联合分布的概率（离散型 r.v.）或密度（连续型 r.v.），称为该样本 $\mathcal{X}$ 的似然值 (likelihood)；似然值与总体分布参数 $\boldsymbol{\theta}$ 间的函数关系，称为似然函数 (likelihood function)，记做 $L(\boldsymbol{\theta}|\mathcal{X})$
- 极大似然估计 (maximum likelihood estimation)，是指给定样本 $\mathcal{X}$ 时，最大化似然函数值

$$\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathcal{X})$$

  的总体参数取值 $\hat{\boldsymbol{\theta}}_N$
- 通常情况下，计算对数似然函数 $\log L(\boldsymbol{\theta}|\mathcal{X})$ 的最大值更为简便，且与最大化似然函数水平值等价 (Why?)

参数估计的基本性质

- 一致性 (consistency)：当样本增加时，参数估计值应该逼近真实值

$$\lim_{N\to\infty} \hat{\theta}_N = \theta_0.$$

通常而言，我们希望上述收敛是以概率 1 收敛，即几乎处处收敛

- 无偏性 (unbiasedness)：参数估计值，作为样本 $\mathcal{X}$ 的函数，在样本联合分布下的期望，等于真实值

$$\mathbb{E}^{\mathbb{P}}\left[\hat{\theta}_N\right] = \theta_0.$$

- 经济、金融中，一致性比无偏性更为重要：前者意味着基于一次抽样所得数据进行的估计，就能够与真实值接近；后者以为这重复很多次抽样所得数据进行估计的平均值，与真实值接近

一致性的基础：大数定律

给定随机变量序列 $\{X_i\}_{i=1}^{\infty}$，一致性估计的理论基础是大数定律 (law of large number)

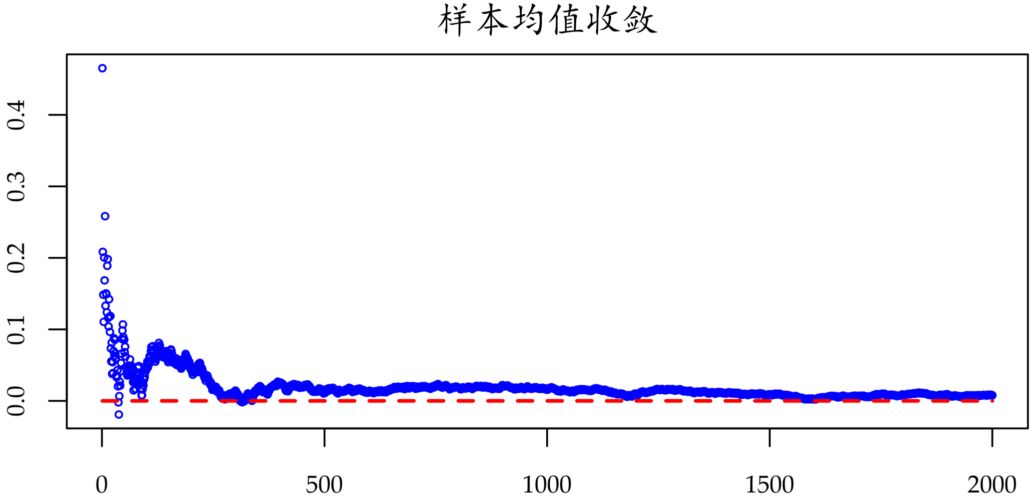
定理 1

当 X_i 满足特定条件时，样本均值以概率 1 收敛到总体均值

$$\lim_{N\to\infty} \hat{\mu}_{X,N} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} X_i = \mathbb{E}X_i$$

记作 $\hat{\mu}_{X,N} \xrightarrow{\text{a.s.}} \mathbb{E}X_i$

示例：$\{X_i\}_{i=1}^{2000} \overset{\text{iid}}{\sim} U([-1,1])$，$\hat{\mu}_K = \sum_{1\le i\le K} X_i$，$\hat{\mu}_K \xrightarrow{\text{a.s.}} 0$



随机变量收敛的概念

收敛：随机变量的收敛概念有 4 种，依分布收敛 $\xrightarrow{d}$，依概率收敛 $\xrightarrow{p}$，矩收敛 $\xrightarrow{m}$，概率-1 收敛（几乎处处收敛）$\xrightarrow{a.s.}$

- 特定条件：弱于 $\{X_i\}$ iid；概率论高级课程中所介绍的“平稳遍历性”条件
- 假设有 r.v. 序列 $\{Z_i\}_{i=1}^{\infty}$ 和 r.v.Z，则称
  1. $Z_i \xrightarrow{d} Z$，若 $\lim_{i\to\infty} F_{Z_i}(z) = F_Z(z)$ 对几乎所有 z 成立，即分布函数逐点收敛
  2. $Z_i \xrightarrow{p} Z$，若 $\forall \varepsilon > 0$，$\lim_{t\to\infty} \mathbb{P}(|Z_i - Z| \geq \varepsilon) = 0$，即 Z_i 与 Z 差别大于任意 ϵ 的概率趋近于 0
  3. $Z_i \xrightarrow{m} Z$，若 $\lim_{i\to\infty} \mathbb{E}[(Z_i - Z)^2] = 0$，差距 2 阶矩收敛到 0
  4. $Z_i \xrightarrow{a.s.} Z$，若 $\mathbb{P}(\lim_{i\to\infty} Z_i = Z) = 1$，$Z_i$ 不收敛到 Z 的样本点集为 0 概率集
- 4 种收敛的关系：$\left\{ \begin{array}{c} \xrightarrow{a.s.} \\ \xrightarrow{m} \end{array} \right\} \Longrightarrow \xrightarrow{p} \Longrightarrow \xrightarrow{d}$
- 本课程中，大数定律中“收敛”，均理解为 $\xrightarrow{a.s.}$

一致性与无偏性示例

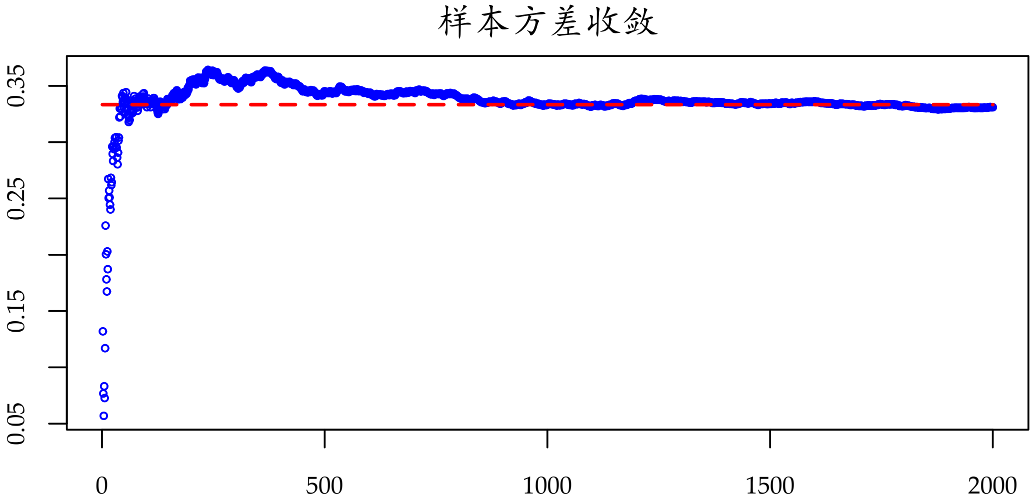
给定 iid 样本 $\{X_i\}_{i=1}^N$，$\mu_X = \mathbb{E}X_i$，$\sigma_X^2 = \mathrm{var}(X_i)$，各阶矩的<u>估计</u> (estimate) 如下

- 期望（均值）：$\hat{\mu}_{X,N} = \frac{1}{N}\sum_i X_i$，一致且无偏
  - $\hat{\mu}_{X,N} \xrightarrow{\text{a.s.}} \mu, \mathbb{E}\hat{\mu}_{X,N} = \mu$
  - 在更一般有序列相关（平稳）条件下，一致性与无偏性依然成立
- 方差：

$$\hat{\sigma}_{X,N}^2 = \begin{cases} \frac{1}{N-1}\sum_i (X_i - \hat{\mu}_X)^2, & \text{iid 下无偏}, \mathbb{E}\hat{\sigma}_{X,N}^2 = \sigma_X^2 \\ \frac{1}{T}\sum_t (X_t - \hat{\mu}_X)^2, & \text{有偏}, \mathbb{E}\hat{\sigma}_{X,N}^2 \neq \sigma_X^2 \end{cases}$$

大样本之下，两个估计值等价，均为一致估计，$\hat{\sigma}_{X,N}^2 \xrightarrow{\text{a.s.}} \sigma_X^2$

示例：$\{X_i\}_{i=1}^{2000} \stackrel{\text{iid}}{\sim} U([-1,1])$，$\hat{\sigma}^2_{X,K} \xrightarrow{\text{a.s.}} \operatorname{var}(X_i) = 1/3$



一致性与无偏性示例

给定 iid 样本 $\{X_i, Y_i\}_{i=1}^N$，$\sigma_{XY} = \mathrm{cov}(X_i, Y_i)$，相关系数为 ρ_{XY}

- 协方差：

$$\hat{\sigma}_{XY,N} = \begin{cases} \frac{1}{N-1}\sum_t (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y), & \text{iid 下无偏,} \mathbb{E}\hat{\sigma}_{XY,N} = \sigma_{XY} \\ \frac{1}{N}\sum_t (X_i - \hat{\mu}_X)(Y_i - \hat{\mu}_Y), & \text{有偏} \mathbb{E}\hat{\sigma}_{XY,N} \neq \sigma_{XY} \end{cases}$$

  大样本之下，两个估计值等价，均为一致估计，$\hat{\sigma}_{XY,N} \xrightarrow{\text{a.s.}} \sigma_{XY}$

- 相关系数：$\hat{\rho}_{XY,N} = \hat{\sigma}_{XY,N}/(\hat{\sigma}_{X,N}\hat{\sigma}_{Y,N})$，一致估计，$\hat{\rho}_{XY,N} \xrightarrow{\text{a.s.}} \rho_{XY}$
  - 可以证明：$|\hat{\rho}_{XY}| \leq 1$

本节内容

1 统计概要

2 常见估计方法

3 统计推断基础

推断的逻辑

- 统计推断和悬疑破案有同样的问题：从已知事实（样本）推断未知真相（参数真实值）
- 真相的可能取值：可能有无穷多种可能，但至少是两种可能

$\Rightarrow$ 即便是无穷多种可能，总可以使用二分法：把可能真相分为互不相交两组 G, B

- 推断的两种结论：支持 G/反对 B vs. 支持 B/反对 G
- 推断有误差 $\Rightarrow$ 4 种结果：

		真相 G	真相 B
推断	支持 G	正确	错误
推断	支持 B	错误	正确

统计假设与两类错误

- 将参数真实值（真相）分为互不相交的两类，等价于做两个相互排斥的假设 (hypothesis)
- 统计学的传统：将最希望被排除的假设，称为原假设 (null hypo.)；另一个假设，称为备择假设 (alternative hypo.)
  - 原假设又称为零假设，一般记做 H_0；备择假设一般记做 H_1 或 H_a
- 假设检验 (hypo. testing)：通过检验决定是否支持假设
  - 第一类错误 (Type-I error)：在 H_0 成立时，支持 H_1
  - 第二类错误 (Type-II error)：在 H_1 成立时，支持 H_0

	H_0 成立	H_1 成立
支持 H_0		第二类错误
支持 H_1	第一类错误	

统计推断的原则：优先控制第一类错误

- 原假设是研究者最希望排除的假设

$\Rightarrow$ 优先控制住拒绝 H_0、接受 H_1 的可能错误：错误拒绝 (false rejection) 或错误否定 (false negative)

  - 如果嫌疑人真是个好人，希望尽可能避免将其判做坏人的情况 $\Rightarrow$ 减少冤假错案
  - 现代统计学诞生于英国，优先控制第一类错误受到英国普通法 (common law) “疑罪从无”原则的影响

- 统计推断就是一个决策过程

	好人	坏人
无罪		逍遥法外 第二类错误
有罪	冤假错案 第一类错误	

统计量、显著性水平与 p-值

- 如何支持/接受或反对/拒绝一个统计假设？
- 检验统计量 (test statistics)：从统计模型出发，从样本随机变量 $\mathcal{X} = \{X_i\}_{i=1}^{N}$ 构造一个统计量 ξ_N (r.v.) 并确定其在特定假设 H_x 下的分布 $\mathbb{P}(\xi_N|H_x)$
- 利用样本实现值，计算统计量取值 $\bar{\xi}_N$；若在 H_x 下观测到 $\bar{\xi}_N$ 的概率 $\mathbb{P}(\xi_N = \bar{\xi}_N|H_x)$ 小于某个临界值，则拒绝该假设
- 该临界值称为该检验的显著性水平 (significance level)：这一概率代表希望控制第一类错误概率的临界值
  - 常用取值包括 1%、5%、10%
- 观测到 $\bar{\xi}_N$ 的概率 $\mathbb{P}(\bar{\xi}_N|H_x)$ 本身，称为检验假设 H_x 的 p-值 (p-value)

检验的功效

- 给定一个检验设计，第一类错误的概率可以表示为

$$\mathbb{P}(\text{拒绝}H_0|H_0\text{为真}) = \mathbb{P}(\text{接受}H_1|H_0\text{为真})$$

第二类错误的概率可以表示为

$$\mathbb{P}(\text{拒绝}H_1|H_1\text{为真}) = \mathbb{P}(\text{接受}H_0|H_1\text{为真})$$

- 第一类错误的概率又称为检验水平 (size of a test)，即原假设下的 p-值
- 检验的功效 (power of a test) 定义为备择假设为真时拒绝原假设的概率，即 $\mathbb{P}(\text{拒绝}H_0|H_1\text{为真})$，故有

$$\text{功效} = \mathbb{P}(\text{拒绝}H_0|H_1\text{为真}) = 1 - \mathbb{P}(\text{接受}H_0|H_1\text{为真}) = 1 - \text{第二类错误概率}$$

- 一个完美的检验是在检验水平很低的同时，功效很大；但通常而言，检验水平与功效间存在矛盾

均值 t-检验：简单情形

- 最常见的假设检验为均值 t-检验，简称 t-检验 (t-test)
- 考虑两个 iid 序列 $\{X_i\}_{i=1}^N$ 与 $\{Y_i\}_{i=1}^N$ 分别服从总体分布 $N(\mu_X, \sigma_X^2)$, $N(\mu_Y, \sigma_Y^2)$
- 简化假设：μ_X, μ_Y 未知，但 σ_X, σ_Y 已知
- 需要检验的原假设为 $H_0 : \mu_X = \mu_Y \Leftrightarrow H_0 : \mu_X - \mu_Y = 0$
- 计算可知，两个样本均值 $\hat{\mu}_X, \hat{\mu}_Y$ 分别服从如下分布

$$\hat{\mu}_X \sim N\left(\mu_X, \tfrac{1}{N}\sigma_X^2\right), \quad \hat{\mu}_Y \sim N\left(\mu_Y, \tfrac{1}{N}\sigma_Y^2\right)$$

- 选取统计量为 $\xi_N = \hat{\mu}_X - \hat{\mu}_Y$，在 H_0 下 ξ_N 的分布为

$$\xi_T = \hat{\mu}_X - \hat{\mu}_Y = N\left(\mu_X - \mu_Y, \tfrac{1}{N}(\sigma_X^2 + \sigma_Y^2)\right) = N\left(0, \tfrac{1}{N}(\sigma_X^2 + \sigma_Y^2)\right)$$

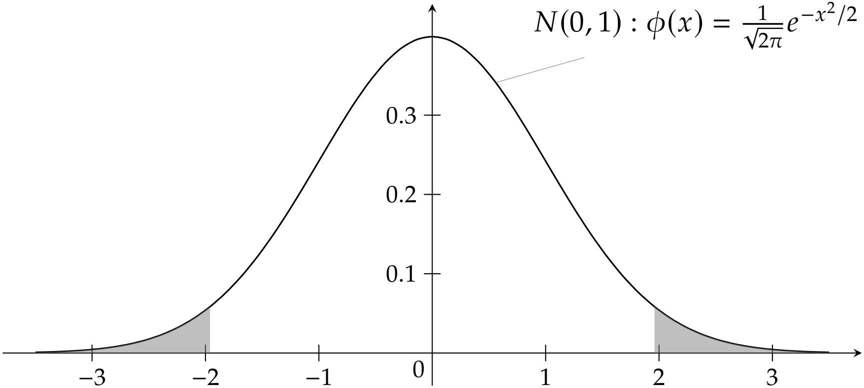
均值 t-检验：简单情形

- 带入样本 $\{X_i, Y_i\}$ 取值，可计算统计量 ξ_N 的取值 $\bar{\xi}_N$
- 考虑如下概率，Φ 为标准正态 cdf：

$$\begin{aligned}\mathbb{P}(|\xi_N| \geq |\bar{\xi}_N|) &= \mathbb{P}\left(\frac{|\xi_N|}{\sigma(\xi_N)} \geq \frac{|\bar{\xi}_N|}{\sigma(\xi_N)}\right) \\ &= 2\Phi\left(-\frac{|\bar{\xi}_N|}{\sigma(\xi_N)}\right)\end{aligned}$$

- 若 $|\bar{\xi}_N|$ 较大，说明观察到 $\xi_N \leq -|\bar{\xi}_N|$ 或 $\xi_N \geq |\bar{\xi}_N|$ 概率很小
- 此时若拒绝 $H_0 : \mu_X = \mu_Y$，出现第一类错误的概率很小！
- $\frac{\xi_N}{\sigma(\xi_N)}$ 称为 <u>t-统计量</u> (t-statistic)

均值 t-检验统计量图示：原假设为正态分布，$\alpha = 5\%$，统计量临界值为 1.96



大样本 t-检验与标准误

- 考虑 iid 样本 $\{X_i\}_{i=1}^N$，服从总体分布 $N(\mu_X, \sigma_X^2)$
- 假设此时 μ_X, σ_X^2 均为未知，并考虑原假设 $H_0 : \mu_X = \mu_0 \equiv 0$
- 自然的想法：检验样本均值 $\hat{\mu}_N$ 是否显著不同于 $\mu_0 = 0$
- H_0 之下，$\hat{\mu}_N = \frac{1}{N}\sum_t X_i$ 的分布为 $N\left(\mu_0, \frac{1}{N}\sigma_X^2\right) = N\left(0, \frac{1}{N}\sigma_X^2\right)$

$$\Rightarrow \frac{\hat{\mu}_N}{\frac{1}{\sqrt{N}}\sigma_X} \sim N(0,1) \Rightarrow \text{统计量}\,\xi_N = \frac{\hat{\mu}_N}{\frac{1}{\sqrt{N}}\hat{\sigma}_{X,N}} \xrightarrow{\mathrm{d}} N(0,1)$$

其中，$\hat{\sigma}_{X,N}^2 = \frac{1}{N}\sum_i (X_i - \hat{\mu}_N)^2$ 为 σ_X^2 的一致估计

- ξ_N 为 $\hat{\mu}_N$ 的标准化 (standardization)，有极限分布 $N(0,1)$
- $\hat{\mu}_N$ 的样本标准差 $\frac{1}{\sqrt{N}}\hat{\sigma}_{X,N}$ 称为标准误 (standard error)