

国际金融试验班 2020 年秋 · 时间序列

# 第 3 讲：统计基础

授课人：刘 岩

武汉大学经管学院金融系

2020 年 9 月 24 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)  $X_1, \dots, X_n$
- 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)  $\mathbf{X}_1, \dots, \mathbf{X}_n$
- 类似的，称这组向量观测值为样本，称  $F$  为总体
- 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  $\mathbf{X}_t = [Y_t, C_t, I_t, G_t, EX_t, IM_t]^T$  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） $\mathcal{X} = \{\mathbf{X}_i\}_{i=1}^n$
- 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  $\{\mathbf{X}_i\}_{i=1}^n$  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  $\{\mathbf{x}_i\}_{i=1}^n$

## 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  $\mathcal{X} = \{\mathbf{X}_i\}_{1 \leq i \leq n}$ ，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  $\{\mathbf{X}_t\}_{t=1}^T$  所满足的统计模型为给定的总体分布

$$M_{\theta} = \mathbb{P}(\{\mathbf{X}_t\}_{t=1}^T | \theta),$$

但总体分布的具体性质未知，表示为参数  $\theta$  真实值  $\theta_0$  未知

- 基本任务：从样本  $\{\mathbf{X}_t\}_{t=1}^T$  推测未知的真实值  $\theta_0$
- 估计：从样本  $\{\mathbf{X}_t\}_{t=1}^T$  获得  $\theta_0$  的估计值  $\hat{\theta}_T$
- 推断：由样本计算得到的  $\hat{\theta}_T$  判断  $\theta$  是否为特定取值  $\theta_0$
- 误差：由于  $\hat{\theta}_T$  继承了样本的随机性，上述推断可能存在误差  $\Rightarrow$  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  $\mathcal{X} = \{X_t\}$  计算样本矩 (sample moment)  $\hat{m}(\mathcal{X})$ ，同时选择总体分布的参数  $\theta$  使得总体矩 (population moment) 与样本矩“**最为**”接近
  - 参数的矩估计值  $\hat{\theta}$  是样本的函数： $\hat{\theta}(\mathcal{X})$ ，常简记为  $\hat{\theta}_T$
- 示例：假设  $\{X_t\}_{t=1}^T \stackrel{\text{iid}}{\sim} U([0, a])$ ， $a$  未知，则可以利用  $\{X_t\}$  的样本均值来估计  $a$ 
  - $X_t$  的总体均值为  $a/2$ ，样本均值为  $\hat{\mu}_T = \frac{1}{T} \sum_t X_t$
  - $a$  的**一个**矩估计为  $\hat{a} = 2\hat{\mu}_T = \frac{2}{T} \sum_t X_t$
  - 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，如最小化平方误差

## 极大似然估计基础

- 假设样本  $\mathcal{X} = \{X_t\}_T$  具有联合分布  $\mathbb{P}(\mathcal{X}|\theta)$
- 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  $\theta$  间的函数关系，称为似然函数 (likelihood function)，记做  $L(\theta|\mathcal{X})$
- 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

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

的总体参数取值  $\hat{\theta}_T$

- 通常情况下，计算对数似然函数  $\log L(\theta|\mathcal{X})$  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

$$\lim_{T \rightarrow \infty} \hat{\theta}_T = \theta_0.$$

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

$$\mathbb{E}^{\mathbb{P}}[\theta_T] = \theta_0.$$

- 经济、金融中, 一致性比无偏性更为重要

## 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：可能有无穷多中可能，但至少是两种可能
- ⇒ 即便是无穷多中可能，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 ⇒ 4 种结果：

		真相	
		$G$	$B$
推断	支持 $G$		
	支持 $B$		

## 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：可能有无穷多中可能，但至少是两种可能
- ⇒ 即便是无穷多中可能，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 ⇒ 4 种结果：

		真相	
		$G$	$B$
推断	支持 $G$	正确	
	支持 $B$		正确

## 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：可能有无穷多中可能，但至少是两种可能
- ⇒ 即便是无穷多中可能，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 ⇒ 4 种结果：

		真相	
		$G$	$B$
推断	支持 $G$	正确	错误
	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 $H_0$		
支持 $H_1$		

## 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 $H_0$		
支持 $H_1$	第一类错误	

## 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
- ⇒ 优先控制住拒绝  $H_0$ 、接受  $H_1$  的可能错误
  - 如果嫌疑人真是个好人的情况 ⇒ 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		
有罪		

## 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
- ⇒ 优先控制住拒绝  $H_0$ 、接受  $H_1$  的可能错误
  - 如果嫌疑人真是个好人，我们希望尽可能避免将其判做坏人的情况 ⇒ 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		
有罪	冤假错案	

## 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
- ⇒ 优先控制住拒绝  $H_0$ 、接受  $H_1$  的可能错误
  - 如果嫌疑人真是个好人，我们希望尽可能避免将其判做坏人的情况 ⇒ 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
有罪	冤假错案	

## 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
- ⇒ 优先控制住拒绝  $H_0$ 、接受  $H_1$  的可能错误
  - 如果嫌疑人真是个好人，我们希望尽可能避免将其判做坏人的情况 ⇒ 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		
有罪	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
- ⇒ 优先控制住拒绝  $H_0$ 、接受  $H_1$  的可能错误
  - 如果嫌疑人真是个好人，我们希望尽可能避免将其判做坏人的情况 ⇒ 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		第二类错误
有罪	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  $\mathcal{X}$  构造一个统计量  $\xi_T$  (r.v.) 并确定其在特定假设  $H_x$  下的分布  $\mathbb{P}(\xi_T|H_x)$
- 利用样本实现值, 计算统计量取值  $\bar{\xi}_T$ ; 若在  $H_x$  下观测到  $\bar{\xi}_T$  的概率  $\mathbb{P}(\xi_T = \bar{\xi}_T|H_x)$  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)
  - 常用取值包括 1%、5%、10%
- 观测到  $\bar{\xi}_T$  的概率  $\mathbb{P}(\bar{\xi}_T|H_x)$  本身, 称为检验假设  $H_x$  的  $p$ -值 ( $p$ -value)

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  $\{X_t\}_{t=1}^T$  与  $\{Y_t\}_{t=1}^T$  分别服从总体分布  $N(\mu_X, \sigma_X^2)$ ,  $N(\mu_Y, \sigma_Y^2)$
- 简化假设： $\mu_X, \mu_Y$  未知，但  $\sigma_X, \sigma_Y$  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

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

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

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

均值  $t$ -检验：简单情形

- 带入样本  $\{X_t, Y_t\}$  取值，可计算统计量  $\xi_T$  的取值  $\bar{\xi}_T$
- 考虑如下概率， $\Phi$  为标准正态 cdf:

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

- 若  $|\bar{\xi}_T|$  较大，说明观察到  $\xi_T \leq -|\bar{\xi}_T|$  或  $\xi_T \geq |\bar{\xi}_T|$  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！
- $\frac{\bar{\xi}_T}{\sigma(\xi_T)}$  称为  $t$ -统计量 ( $t$ -statistic)

大样本  $t$ -检验与标准误

- 考虑 iid 样本  $\{X_t\}_{t=1}^T$ , 服从总体分布  $N(\mu_X, \sigma_X^2)$
- 假设此时  $\mu_X, \sigma_X^2$  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  $\hat{\mu}_T$  是否显著不同于  $\mu_0 = 0$
- $H_0$  之下,  $\hat{\mu}_T = \frac{1}{T} \sum_t X_t$  的分布为  $N(\mu_0, \frac{1}{T} \sigma_X^2) = N(0, \frac{1}{T} \sigma_X^2)$

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

其中,  $\hat{\sigma}_{X,T}^2 = \frac{1}{T} \sum_t (X_t - \hat{\mu}_T)^2$  为  $\sigma_X^2$  的一致估计

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