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

第六章 统计抽样理论导论

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

目　录

- **统计分析**是建立在大量相同或相似重复**试验**之结果的基础上。
- 假设**随机变量** X_i 表示第 i 次试验的结果。
- 若进行 n 次试验，可获得试验的结果序列 $X_1,\cdots,X_n$。
  - ✓ n 被称作**样本容量**，
  - ✓ 结果序列 $X_1,\cdots,X_n$ 构成了一个**随机样本**。
- 基于样本信息，可推断生成该观测结果序列的**概率分布**。

定义 6.1

[随机样本 (random sample)]：一个随机样本是由 n 个随机变量 $X_1,\cdots,X_n$ 所构成的序列，记作 $\boldsymbol{X}^n=(X_1,\cdots,X_n)$。

- 随机样本 $\boldsymbol{X}^n$ 的一个**实现值** (realization) 称为从随机样本 $\boldsymbol{X}^n$ 生成的一个**数据集** (data set) 或**样本点** (sample point)，$\boldsymbol{X}^n$ 的一个样本点记作 $\boldsymbol{x}^n=(x_1,\cdots,x_n)$。
- 一个随机样本 $\boldsymbol{X}^n$ 可生成多个不同的数据集。所有可能的 $\boldsymbol{X}^n$ 的样本点构成随机样本 $\boldsymbol{X}^n$ 的**样本空间** (sample space)。

例 6.1：[抛硬币]

- 假设抛掷 n 枚硬币。令 X_i 表示抛掷第 i 枚硬币的结果，
  - ✓ 若正面朝上则 $X_i = 1$，
  - ✓ 否则 $X_i = 0$。
- 那么 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 构成一个随机样本。
  - ✓ 抛掷 n 枚硬币后，将获一个实数序列，如

$$\boldsymbol{x}^n = (1,1,0,0,1,0,\cdots,1)$$

该序列是来自随机样本 $\boldsymbol{X}^n$，样本容量为 n 的一个**数据集**，也称为 $\boldsymbol{X}^n$ 的一个**样本点**。

例 6.1 (Cont.):

✓ 若再次抛掷这 n 枚硬币，可获得一个不同实数序列，如

$$\boldsymbol{x}^n = (1,0,0,1,1,1,\cdots,0)$$

这是来自随机样本 $\boldsymbol{X}^n$ 的**另一数据集**或**样本点**。

- 随机样本 $\boldsymbol{X}^n$ 一共可生成 $\boldsymbol{2^n}$ **个数据集**，每个数据集的**样本容量均为 $\boldsymbol{n}$**。

例 6.2： [中国 GDP 年增长率]

- 令 X_i 表示 1953 至 2010 年间第 i 年的中国 GDP 年增长率，则 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 构成了一个样本容量 $n = 58$ 的随机样本。图 6.1 所示的观测数据 $\boldsymbol{x}^n = (x_1, \cdots, x_n)$ 是 $\boldsymbol{X}^n$ 的一个实现值。

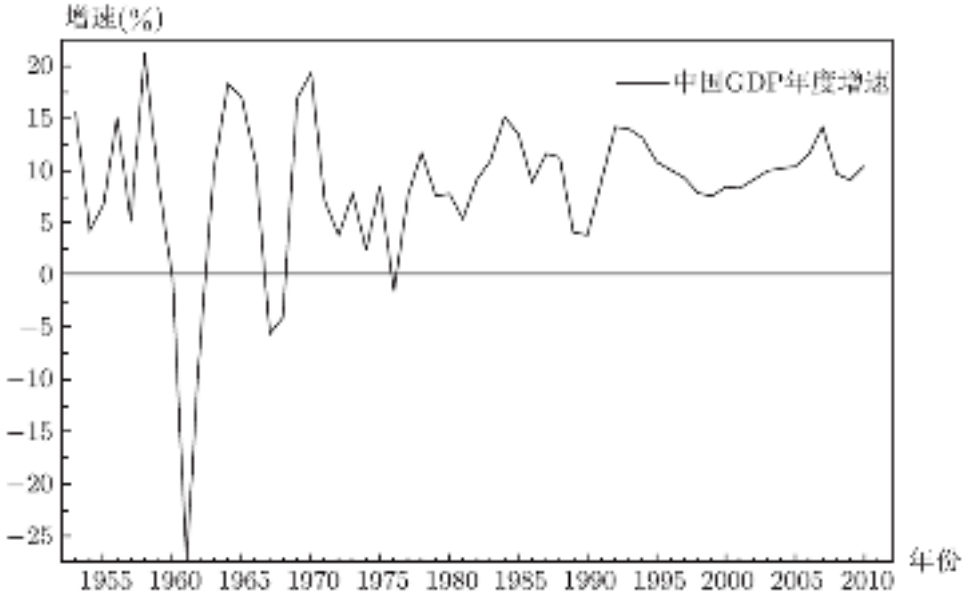

图 6.1：中国 GDP 年增长率

现实中常常只能观测或获得一个数据集 x^n

- 尽管在理论上，一个随机样本 $\boldsymbol{X}^n$ 可生成许多不同的样本容量均为 n 的数据集 $\boldsymbol{x}^n$，但在现实中常常只能观测或获得一个数据集 $\boldsymbol{x}^n$，就像例 6.2 和例 6.3 那样。
- 例如，若希望获得中国 GDP 年增长率的另一个数据集 (即另一个不同的实现值序列)，则必须让中国经济回到 1953 年后重新运行。由于现实经济具有非试验性，这显然是无法实现的。有时，即便可重复某些社会经济实验，其成本也可能过于高昂而不能实现。
- 然而，在统计分析中，我们仍然假设例 6.2 或例 6.3 中的观测数据是随机样本 $\boldsymbol{X}^n$ 众多可能的实现值之一。

随机变量序列 $X_1,\cdots,X_n$ 及其实现值的顺序可能是不可随意改变的

- 例 6.2 的时间序列随机样本即是一个例子，其随机变量 $X_1,\cdots,X_n$ 并非互相独立，且 X_i 可能依赖于之前的经济增长率 $\{X_{i-1},X_{i-2},\cdots\}$。若改变随机变量序列及其实现值的顺序，则它们的动态结构信息将无法保存。

例 6.3：

- 令 X_i 为 1960 年 1 月 4 日至 2010 年 12 月 31 日之间第 i 日的 S&P 500 价格指数日收益率，则 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 构成了一个样本容量为 $n = 12839$ 的随机样本。图 6.2 所示的观测数据集 $\boldsymbol{x}^n = (x_1, \cdots, x_n)$ 是随机样本 $\boldsymbol{X}^n$ 的一个实现值。

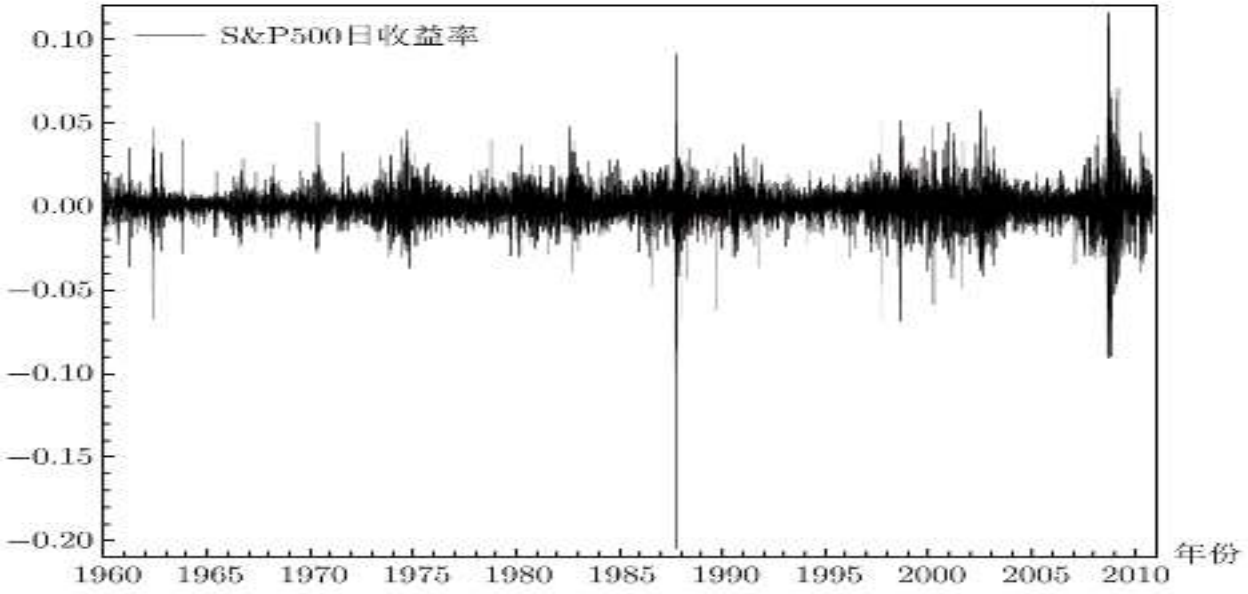

图 6.2： S&P 500 收盘价格指数日收益率

n 个随机变量的联合 PMF/PDF

- 一个随机样本 $\boldsymbol{X}^n$ 可视为一个 n 维随机向量，即 $\boldsymbol{X}^n: S \to \mathbb{R}^n$，其中 S 为随机试验的样本空间。随机样本 $\boldsymbol{X}^n$ 的信息可由 n 个随机变量的联合 PMF/PDF 完整描述，即

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^{\mathrm{n}}) = \prod_{i=1}^{n} f_{X_i|\boldsymbol{X}^{i-1}}\left(x_i \mid \boldsymbol{x}^{i-1}\right)$$

- 按照惯例，通常记 $f_{X_1|\boldsymbol{X}^0}\left(x_1 \mid \boldsymbol{x}^0\right) = f_{X_1}(x_1)$，并称之为随机变量 X_1 的边际 PMF/PDF。
- 这里，条件 PMF/PDF 的乘积系通过对联合概率反复应用**概率乘法法则**而得。联合 PMF/PDF 可用于计算随机样本 $\boldsymbol{X}^n$ 及其各种函数的联合概率。

独立样本与时间序列样本

- 上述随机样本的定义涵盖了独立样本和时间序列样本。
  - ✓ 对**独立样本** (independent samples) 而言，随机样本中的随机变量 $X_1,\cdots,X_n$ 互相独立；
  - ✓ 对**时间序列样本** (time series samples) 而言，随机样本中的随机变量 $X_1,\cdots,X_n$ 并非相互独立。
- 为聚焦于统计学的基本思想，本书将主要考虑独立随机样本的情形。

定义 6.2

[IID 随机样本 (IID Random Sample)]：若

(1) 随机变量 $X_1,\cdots,X_n$ **相互独立**；

(2) 每个随机变量 X_i 具有**相同的边际分布** $F_X(x)$，

则称随机变量序列 $X_1,\cdots,X_n$ 为来自总体分布为 $F_X(x)$，样本容量为 n 的**独立同分布 (IID)** 随机样本。

◆ 问题 6.1

如何解释 IID 随机样本？其含义是什么？

- **同分布**意味着同类试验重复进行，**独立性**则说明试验之间是独立进行的，因而每次试验都可获得新的信息（若 $X_1,\cdots,X_n$ 高度相关，其实现值将非常相似且变化很少，故包含的新信息也较少)。
- **统计分析的主要目的**：基于大量重复同类试验所生成的**观测数据推断总体分布** $F_X(x)$。

参数建模方法

- 在很多实际应用中，通常假设相对应的 PMF/PDF 为 $f_X(x) = f(x,\theta)$，其中函数形式 $f(\cdot,\cdot)$ 已知，但参数 θ 的值未知。
- 例如，假设 $\boldsymbol{X}^n$ 是来自总体分布为 $N(\mu,\sigma^2)$ 的随机样本，则

$$f_X(x) = f(x,\theta)$$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, -\infty < x < \infty$$

其中 $\theta = (\mu,\sigma^2)$。

- 给定来自随机样本 $\boldsymbol{X}^n$ 的观测数据 $\boldsymbol{x}^n$，可推断参数 $\theta = (\mu,\sigma^2)$ 的真实值。

基于随机样本推断总体是统计分析最重要的特征

◆ 问题

(1) 对于随机样本有何要求?

(2) 给定一个随机样本，什么是最好的推断方法?

(3) 如果随机样本有一些缺陷（样本选择偏差、数据缺失、异常值等）应该怎么办?

(4) 给定一个数据集 $\boldsymbol{x}^n$，如何从中提取有用信息？什么分析工具可用于实现该目标?

定义 6.3

[统计量 (Statistic)]：令 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自某一总体，样本容量为 n 的随机样本。统计量 $T(\boldsymbol{X}^n) = T(X_1, \cdots, X_n)$ 是随机样本 $\boldsymbol{X}^n$ 的**实值或向量值函数**。

- 函数 $T(\cdot)$ 是从 n 维样本空间 $\boldsymbol{X}^n$ 到**低维**欧氏空间的一个映射。简便起见，这里省略了函数形式 $T(\cdot)$ 对样本容量 n 可能存在的依赖。
- 统计量 $T(\boldsymbol{X}^n)$ **不包含任何未知参数**，它完全是随机样本 $\boldsymbol{X}^n$ 的函数。给定任何一个数据集 $\boldsymbol{x}^n$，可获得统计量 $T(\boldsymbol{X}^n)$ 的一个实数值或向量值。

统计量的用途

- 统计量 $T(\boldsymbol{X}^n)$ 可用于：
  - ✓ **有效刻画数据的某些特征**，如：
    - ➢ 最大值、最小值、中位数、均值、标准差等
  - ✓ **估计未知参数值**
  - ✓ 进行**参数假设检验**等
- 统计量的**可解释性**非常重要。

例 6.4

- 令 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为一个随机样本，样本均值和样本方差为两个经典统计量。
  - ✓ **样本均值**

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

  - ✓ **样本方差**

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

◆ 问题

- 样本均值 $\bar{X}_n$ 和样本方差 S_n^2 可用于估计总体分布 $F_X(x)$ 的均值 μ_X 和方差 σ_X^2。为什么 $\bar{X}_n$ 和 S_n^2 分别是 μ_X 和 σ_X^2 的“良好” 估计量?

- 第七章和第八章将更正式地引入各种收敛概念度量估计量和待估参数值之间的接近程度。

例 6.5

- 令 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自总体 $f(x, \theta)$ 的 IID 随机样本，其中 θ 是未知参数。则 $\boldsymbol{X}^n$ 的联合 PMF/PDF 的对数

$$\hat{L}(\theta \mid \boldsymbol{X}^n) = \ln \prod_{i=1}^{n} f(X_i, \theta) = \sum_{i=1}^{n} \ln f(X_i, \theta)$$

  称为随机样本 $\boldsymbol{X}^n$ 关于参数 θ 的**对数似然函数**。

- $\hat{L}(\theta \mid \boldsymbol{X}^n)$ 依赖于随机样本 $\boldsymbol{X}^n$，但它不是一个统计量，因为它还依赖于未知参数 θ。

定义 6.4

[抽样分布 (Sampling Distribution)]：统计量 $T(\boldsymbol{X}^n)$ 的概率分布称为 $T(\boldsymbol{X}^n)$ 的抽样分布。

- 因为 $T(\boldsymbol{X}^n)$ 是 n 个随机变量的函数， $T(\boldsymbol{X}^n)$ 本身是一个**随机变量**或**随机向量**。
- 由于该分布通常可由随机样本中随机变量 $X_1,\cdots,X_n$ 的联合分布推导而来，所以 $T(\boldsymbol{X}^n)$ 的分布称为**抽样分布**。
- $T(\boldsymbol{X}^n)$ 的抽样分布不同于总体分布 $F_X(x)$，后者是 IID 随机样本 $\boldsymbol{X}^n$ 中的每个随机变量 X_i 的边际分布。
- 若 $\boldsymbol{X}^n$ 是 IID 随机样本，则推导很简单。统计量 $T(\boldsymbol{X}^n)$ 的抽样分布在统计推断中扮演非常重要的角色。

目　录

定义 6.5

[样本均值 (Sample Mean)]：设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自均值为 μ，方差为 σ^2 的总体分布的一个随机样本，则

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

称为随机样本 $\boldsymbol{X}^n$ 的样本均值。

- 因为 $X_1, \cdots, X_n$ 是随机变量，故 $\bar{X}_n$ 也是**随机变量**。$\bar{X}_n$ 的分布称为 $\bar{X}_n$ 的**抽样分布**。当仅有一个观测样本点 $\boldsymbol{x}^n$ 时，样本均值 $\bar{x}_n$ 看似不是随机的。然而，若注意到观测样本 $\boldsymbol{x}^n$ 是许多可能被抽取的样本点的其中之一，且每个观测样本一般会有不同的样本均值，那么就会理解样本均值事实上是随机的。

定理 6.1

假设 $\boldsymbol{X}^n$ 为一个随机样本，则

$$\bar{X}_n = \arg \min_{-\infty<a<\infty} \sum_{i=1}^{n} (X_i - a)^2$$

- 目标函数 $\sum_{i=1}^{n} (X_i - a)^2$ 称为**残差平方和** (sum of squared residuals, SSR)。
- 该定理表明，样本均值 $\bar{X}_n$ 是**最小化残差平方和** $\sum_{i=1}^{n} (X_i - a)^2$ 的**最优解**。
- 实际上，样本均值是如下线性回归模型的**普通最小二乘 (ordinary least squares, OLS) 估计量**：$X_i = a + \varepsilon_i$，其中 $\{\varepsilon_i\}$ 为 $E(\varepsilon_i) = 0$, $\text{var}(\varepsilon_i) = \sigma^2$ 的 IID 随机扰动项序列。

- 现在考察样本均值 $\bar{X}_n$ 的**概率统计性质**，这对推断未知总体均值 μ 非常重要。
- 特别地，需探讨如下几个问题：
  - ✓ $\bar{X}_n$ 的**均值**
  - ✓ $\bar{X}_n$ 的**方差**
  - ✓ $\bar{X}_n$ 的**抽样分布**

定理 6.2

假设 $X_1, \cdots, X_n$ 为具有相同总体均值 μ 的 n 个同分布随机变量序列，则对所有 $n \geq 1$，

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

证明：

$$\begin{aligned} E(\bar{X}_n) &= \frac{1}{n}\sum_{i=1}^{n} E(X_i) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mu \\ &= \mu \end{aligned}$$

证毕。

样本均值的期望等于总体均值的含义

- 对任意给定样本容量 $n \geq 1$，样本均值 $\bar{X}_n$ 的期望等于总体均值 μ。这一结果不要求随机变量 $X_1, \cdots, X_n$ 相互独立。
- 直觉上，$E(\bar{X}_n) = \mu$ 表明样本均值 $\bar{X}_n$ 在估计总体均值 μ 时不存在系统性误差。若生成大量样本容量为 n 的数据集 $\boldsymbol{x}^n$，每个数据集提供 $\bar{X}_n$ 的一个实现值 $\bar{x}_n$，则对任意给定样本容量 n，这些样本均值的平均将充分接近总体均值 μ，即不存在系统性向上或向下偏离总体均值 μ 的误差（**无系统性偏差**）。

定理 6.3

假设 $\boldsymbol{X}^n$ 是来自均值为 μ，方差为 σ^2 的 IID 随机样本。则对所有$n \geq 1$，

$$\mathrm{var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

证明：

- 当 X 和 Y 相互独立时，有

$$\begin{aligned}\mathrm{var}(a + bX + cY) &= b^2\sigma_X^2 + c^2\sigma_Y^2 + 2bc\,\mathrm{cov}(X, Y) \\ &= b^2\sigma_X^2 + c^2\sigma_Y^2\end{aligned}$$

证明(Cont.):

- 类似地，对于 IID 随机样本 $\boldsymbol{X}^n$，有

$$\begin{aligned}\operatorname{var}(\bar{X}_n) &= \operatorname{var}\left(n^{-1}\sum_{i=1}^{n} X_i\right) \\ &= n^{-2}\sum_{i=1}^{n}\operatorname{var}(X_i) \\ &= n^{-2}\sum_{i=1}^{n}\sigma^2 \\ &= \frac{\sigma^2}{n}\end{aligned}$$

- **证毕。**

定理 6.3 的含义

- 与每个随机变量 X_i 的总体方差 σ^2 不同，$\bar{X}_n$ 的方差 σ^2/n 测度了样本均值 $\bar{X}_n$ 距离其中心 $E(\bar{X}_n)$ 的远近。
- $\mathrm{var}(\bar{X}_n)=\sigma^2/n$ 表明 $\bar{X}_n$ 对其中心 $E(\bar{X}_n)$ 的偏离程度随样本容量 $n \to \infty$ 而趋于 0。
- 由于 $E(\bar{X}_n) = \mu$，当 $n \to \infty$ 时，$\bar{X}_n$ 的**均方误 (MSE)** 为

$$E(\bar{X}_n - \mu)^2 = \mathrm{var}(\bar{X}_n) = \frac{\sigma^2}{n} \to 0$$

- $\mathrm{var}(\bar{X}_n) = \sigma^2/n$ 在统计学上看似简单，却是金融学中风险分散化 (risk diversification) 原理的依据所在。

例 6.6：[通过分散化消除特质性风险 (Idiosyncratic Risk Elimination via Diversification)]

- 根据经典**资本资产定价模型** (CAPM)，资产 i 在一定持有期内的收益率可表示为

$$R_i = \alpha + \beta_i R_m + \varepsilon_i$$

- 其中
  - ✓ α 是一个常数，代表**无风险资产收益率**，
  - ✓ R_m 是所有资产都面临的**共同的市场风险因子** (一般以市场组合收益率表示)，
  - ✓ β_i 是**因子载荷系数** (factor loading coefficient) 或 **beta 系数**，
  - ✓ ε_i 则代表资产 i 的**特质性风险**。

例 6.6 (Cont.):

- 进一步假设 $\varepsilon_1,\cdots,\varepsilon_n$ 是均值为 0，方差为 σ^2 的独立同分布序列，且与市场风险因子 R_m 不相关。
- 资产 i 的风险用其方差测度，等于

$$\mathrm{var}(R_i) = \beta_i^2 \mathrm{var}(R_m) + \sigma^2$$

- 其中
  - ✓ $\beta_i^2 \mathrm{var}(R_m)$ 是无法避免的**市场系统性风险**，
  - ✓ 而资产 i 的**特质性风险** σ^2 则可通过构建一个包含大量不同资产的投资组合加以消除。

例 6.6 (Cont.):

- 为说明这一点，考察 n **个资产的等权重投资组合之收益率**

$$\bar{R}_n = \sum_{i=1}^{n} \frac{1}{n} R_i = \alpha + \bar{\beta}_n R_m + \bar{\varepsilon}_n$$

- 其中，当 $n \to \infty$ 时
  - ✓ $\bar{\beta}_n = n^{-1} \sum_{i=1}^{n} \beta_i \to \beta \neq 0$，
  - ✓ $\bar{\varepsilon}_n = n^{-1} \sum_{i=1}^{n} \varepsilon_i$ 为 n 个资产的特质性风险样本 $(\varepsilon_1, \cdots, \varepsilon_n)$ 的样本均值。
- 因为 $\mathrm{var}(\bar{\varepsilon}_n) = \sigma^2/n$，则当 $n \to \infty$，有

$$\mathrm{var}(\bar{R}_n) = \bar{\beta}_n^2 \mathrm{var}(R_m) + \frac{\sigma^2}{n} \to \beta^2 \mathrm{var}(R_m)$$

- 因此，与单个资产有关的特质性风险可通过在投资组合中纳入足够多数目的不同资产加以消除。

定理 6.4

假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为 IID 正态分布随机样本，其中总体均值为 μ，总体方差为 $\sigma^2 < \infty$。定义**标准化样本均值** (standardized sample mean)

$$Z_n = \frac{\bar{X}_n - E(\bar{X}_n)}{\sqrt{\mathrm{var}(\bar{X}_n)}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}$$

则对所有 $n \geq 1$，

$$Z_n \sim N(0,1)$$

证明：

- 令 $Y_i = (X_i - \mu)/\sigma$，则 $Y_i \sim N(0,1)$。由定理 4.1 可得 Y_i 的 MGF 为

$$M_{Y_i}(t) = e^{\frac{1}{2}t^2}, \text{对所有 } t \text{ 值}$$

证明 (Cont.):

- 现在考察 $Z_n = n^{-1/2}\sum_{i=1}^{n} Y_i$ 的 MGF：

$$
\begin{aligned}
M_{Z_n}(t) &= E\left(e^{tZ_n}\right) \\
&= E\left(e^{tn^{-\frac{1}{2}}\sum_{i=1}^{n} Y_i}\right) \\
&= E\left(\prod_{i=1}^{n} e^{tn^{-\frac{1}{2}}Y_i}\right) \\
&= \prod_{i=1}^{n} E\left(e^{tn^{-\frac{1}{2}}Y_i}\right) \\
&= \prod_{i=1}^{n} M_{Y_i}\left(tn^{-\frac{1}{2}}\right) \\
&= \left[e^{\frac{1}{2}\left(tn^{-\frac{1}{2}}\right)^2}\right]^n \\
&= e^{\frac{1}{2}t^2}
\end{aligned}
$$

- 则对所有 $n \geq 1$ 有 $Z_n \sim N(0,1)$。

证毕。

目　录

- 上一节介绍如何用样本均值 $\bar{X}_n$ 估计总体均值 μ，并证明 $\bar{X}_n$ 是 μ 的良好估计量，即当 $n \to \infty$ 时，$E(\bar{X}_n - \mu)^2 = \sigma^2/n \to 0$。

◆ **问题**

这一节的问题是，如何估计总体方差 σ^2?

- 回顾方差公式

$$\sigma^2 = E(X_i - \mu)^2$$

- σ^2 的一个可能估计量为

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$$

- 但若 μ 未知，则该估计量仍然未知。若用样本均值 $\bar{X}_n$ 代替 μ，则 $(X_i - \bar{X}_n)^2$ 的均值为

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

- 事实上，通常使用如下的样本方差估计量

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

- 之所以除以因子 $n-1$ 而不是 n，是因为这里用样本均值 $\bar{X}_n$ 替代未知总体均值 μ。

- 样本方差估计量 S_n^2 是随机样本 $\boldsymbol{X}^n$ 的非线性函数。
- 与研究样本均值 $\bar{X}_n$ 的性质一样，需要考察 S_n^2 的如下性质：
  - ✓ S_n^2 的**均值**；
  - ✓ S_n^2 的**方差**；
  - ✓ S_n^2 的**抽样分布**。
- S_n^2 的这些概率统计性质在涉及 S_n^2 的统计推断中十分重要。

定理 6.5

假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自总体 (μ, σ^2) 的 IID 随机样本。则对于所有 $n > 1$，$E\left(S_n^2\right) = \sigma^2$。

证明：

- 根据公式$(a-b)^2 = a^2 - 2ab + b^2$，有

$$\begin{aligned}
\sum_{i=1}^n (X_i - \bar{X}_n)^2 &= \sum_{i=1}^n [(X_i - \mu) - (\bar{X}_n - \mu)]^2 \\
&= \sum_{i=1}^n (X_i - \mu)^2 - 2\sum_{i=1}^n (X_i - \mu)(\bar{X}_n - \mu) + \sum_{i=1}^n (\bar{X}_n - \mu)^2 \\
&= \sum_{i=1}^n (X_i - \mu)^2 - 2(\bar{X}_n - \mu)\sum_{i=1}^n (X_i - \mu) + n(\bar{X}_n - \mu)^2 \\
&= \sum_{i=1}^n (X_i - \mu)^2 - 2n(\bar{X}_n - \mu)^2 + n(\bar{X}_n - \mu)^2 \\
&= \sum_{i=1}^n (X_i - \mu)^2 - n(\bar{X}_n - \mu)^2
\end{aligned}$$

证明 (Cont.):

- 两边取期望得

$$E\sum_{i=1}^{n}(X_i-\bar{X}_n)^2=\sum_{i=1}^{n}E(X_i-\mu)^2-nE[(\bar{X}_n-\mu)^2]$$
$$=n\sigma^2-n\frac{\sigma^2}{n}$$
$$=(n-1)\sigma^2$$

- 上述推导中，用到了 $\sum_{i=1}^{n}(X_i-\mu)=n(\bar{X}_n-\mu)$，以及第六章第二节的 $E(\bar{X}_n-\mu)^2=\frac{\sigma^2}{n}$。

- 因此

$$E\left(S_n^2\right)=E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2\right]=\sigma^2$$

证毕。

- 上述推导过程中用到了 $E(\bar{X}_n - \mu)^2 = \sigma^2/n$，故假设 n 个随机变量 $X_1, \cdots, X_n$ 相互独立非常重要。
- 如前所述，使用因子 $n-1$ 而非 n 旨在确保 S_n^2 为 σ^2 的无偏估计量 (unbiased estimator)，即 $E(S_n^2) = \sigma^2$。
- 因为 μ 未知，需用 $\bar{X}_n$ 来替代。这导致损失了一个用于计算样本方差 S_n^2 的观测值，使数据集的自由度从 n 减少到 $n-1$。

- 在第四章中，若非负随机变量服从 χ^2 分布，则其 PDF 为

$$f_X(x)=\frac{1}{\sqrt{2^{\nu}}\Gamma\left(\frac{\nu}{2}\right)}x^{\frac{\nu}{2}-1}e^{-\frac{x}{2}},x>0$$

  ✓ 其中 ν 称为卡方分布的自由度。

- χ^2_{ν} 的自由度 ν 不必为整数。但当 ν 为整数时，随机变量 χ^2_{ν} 有更为直观的表示。

引理 6.1

[χ^2 分布]：令 $Z_1,\cdots,Z_\nu$ 为 IID $N(0,1)$ 随机变量，其中 ν 为正整数。则

$$\sum_{i=1}^{\nu} Z_i^2 \sim \chi_\nu^2$$

即 ν 个相互独立的 $N(0,1)$ 随机变量的平方和服从 χ_ν^2 分布。

证明：

- 当 $Z_i \sim N(0,1)$ 时，有 $Z_i^2 \sim \chi_1^2$，其 MGF 为

$$M_{Z_i^2}(t) = (1-2t)^{-\frac{1}{2}}, t < \frac{1}{2}$$

证明 (Cont.):

- 令 $X = \sum_{i=1}^{\nu} Z_i^2$。则给定 $Z_1, \cdots, Z_\nu$ 相互独立，有

$$\begin{aligned} M_X(t) &= E\left(e^{tX}\right) \\ &= E\left(e^{t\sum_{i=1}^{\nu} Z_i^2}\right) \\ &= \prod_{i=1}^{\nu} E\left(e^{tZ_i^2}\right) \\ &= \left[(1-2t)^{-\frac{1}{2}}\right]^{\nu} \\ &= (1-2t)^{-\frac{\nu}{2}} \end{aligned}$$

- 根据 MGF 的唯一性，有 $X \sim \chi_\nu^2$。这称为 χ^2 分布的可加性。

证毕。

χ_ν^2 分布

- χ_ν^2 分布是伽玛分布的一个特例，即 $\text{Gamma}\left(\frac{\nu}{2}, 2\right)$，其均值和方差分别为

$$E\left(\chi_\nu^2\right) = \nu$$

和

$$\text{var}\left(\chi_\nu^2\right) = 2\nu$$

- χ_ν^2 形状不关于均值对称，而是偏向右边。当 $\nu \to \infty$ 时，偏度不断减小至零。

定理 6.6

假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为 IID $N(\mu, \sigma^2)$ 随机样本。则对每个 $n > 1$

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

其中 χ_{n-1}^2 是自由度为 $n-1$ 的卡方分布。

证明：

- 容易建立以下**递归关系 (recursive relation)**

$$(n-1)S_n^2 = (n-2)S_{n-1}^2 + \frac{n-1}{n}(X_n - \bar{X}_{n-1})^2$$

- 现用**归纳法 (induction)** 证明该定理，包含下述两个步骤。

证明 (Cont.):

- **步骤一：** 首先考察 $n=2$ 的情况。因为 $(X_2-X_1)/\sqrt{2}\sigma \sim N(0,1)$，故有

$$\begin{aligned}\frac{(2-1)S_2^2}{\sigma^2} &= \frac{1}{2\sigma^2}(X_2-X_1)^2 \\ &= \left(\frac{X_2-X_1}{\sqrt{2}\sigma}\right)^2 \\ &\sim \chi_1^2\end{aligned}$$

证明 (Cont.):

- **步骤二：**假设对任意正整数 $n = \nu > 1$，有 $(\nu - 1)S_\nu^2/\sigma^2 \sim \chi_{\nu-1}^2$。现在证明 $\nu S_{\nu+1}^2/\sigma^2 \sim \chi_\nu^2$。
- 对 $n = \nu + 1$，有

$$\frac{\nu S_{\nu+1}^2}{\sigma^2} = \frac{(\nu - 1)S_\nu^2}{\sigma^2} + \frac{\nu}{(\nu + 1)\sigma^2}(X_{\nu+1} - \bar{X}_\nu)^2$$

- 此处 $X_{\nu+1} \sim N(\mu, \sigma^2), \bar{X}_\nu \sim N\left(\mu, \frac{1}{\nu}\sigma^2\right)$，且 $X_{\nu+1}$ 和 $\bar{X}_\nu$ 相互独立。

证明 (Cont.):

- 因此

$$X_{\nu+1} - \bar{X}_{\nu} \sim N\left(0, \sigma^2 + \frac{\sigma^2}{\nu}\right)$$

或等价地

$$\sqrt{\frac{\nu}{(\nu+1)\sigma^2}}(X_{\nu+1} - \bar{X}_{\nu}) \sim N(0,1)$$

- 故有 $\frac{\nu}{\nu+1}(X_{\nu+1} - \bar{X}_{\nu})^2/\sigma^2 \sim \chi_1^2$。

证明 (Cont.)：

- 如果这一项和 S_v^2 相互独立，则给定 $(v-1)S_v^2/\sigma^2 \sim \chi_{v-1}^2$ 且两个相互独立的 χ^2 随机变量之和服从 χ^2 分布(关于 χ^2 分布的可加性，参见例 5.33)，可得 $vS_{v+1}^2/\sigma^2 \sim \chi_v^2$。
- 因此，若能证明以下 S_n^2 和 $\bar{X}_n$ 相互独立的结论，则该定理得证。

证毕。

定理 6.7

设 $\boldsymbol{X}^n$ 为 IID $N(\mu,\sigma^2)$ 随机样本，则对任意 $n>1$，S_n^2 和 $\bar{X}_n$ 相互独立。

- 尽管 S_n^2 和 $\bar{X}_n$ 均为相同随机变量 $\{X_i\}_{i=1}^n$ 的函数，但定理 6.7 表明它们相互独立。
- S_n^2 和 $\bar{X}_n$ 相互独立来自随机样本 $\boldsymbol{X}^n$ 的正态性假设 (参考第五章例 5.21)。为证明定理 6.7，可使用下述引理。

引理 6.2

假设 $X_j \sim \text{IID}\ N(\mu, \sigma^2), j = 1, \cdots, n$。对常数 a_{ij} 和 b_{rj}，定义

$$U_i = \sum_{j=1}^{n} a_{ij} X_j, i = 1, \cdots, v$$

$$V_r = \sum_{j=1}^{n} b_{rj} X_j, r = 1, \cdots, m$$

其中 $v + m \leq n$，则

(1) 对任何一对 (i, r)，当且仅当 $\text{cov}(U_i, V_r) = 0$ 时，随机变量 U_i 和 V_r 相互独立；

(2) 当且仅当对所有 $i \in \{1, \cdots, v\}$ 和 $r \in \{1, \cdots, m\}$，U_i 和 V_r 相互独立时，随机向量 $(U_1, \cdots, U_v)$ 和 $(V_1, \cdots, V_m)$ 相互独立。

- 直觉上，所有的 U_i 和 V_r 服从联合正态分布。因此，对所有 i,r，当且仅当每一对随机变量 U_i 和 V_r 的协方差为零时，二者相互独立。

证明 (定理 6.7):

- 由于 $S_n^2 = (n-1)^{-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$ 是 n 个随机变量 $(X_1 - \bar{X}_n), \cdots, (X_n - \bar{X}_n)$ 的函数，只需证明 $\bar{X}_n$ 和 $\{(X_1 - \bar{X}_n), \cdots, (X_n - \bar{X}_n)\}$ 相互独立即可。
- 利用引理6.2。令 $U_1 = \bar{X}_n - \mu$，$V_r = X_r - \bar{X}_n$。首先证明对所有 $r = 1, \cdots, n$，U_1 和 V_r 相互独立。

证明 (Cont.)：

- 因为对任意给定 $r \in \{1, \cdots, n\}$，有

$$\begin{aligned}\operatorname{cov}(U_1, V_r) &= E(U_1 V_r) \\ &= E[(\bar{X}_n - \mu)(X_r - \mu)] - E(\bar{X}_n - \mu)^2 \\ &= \frac{\sigma^2}{n} - \frac{\sigma^2}{n} \\ &= 0\end{aligned}$$

- 根据引理 6.2 (1)，有 U_1 和 V_r 相互独立。再由引理 6.2 (2) 即得 U_1 和 $(V_1, \cdots, V_n)$ 相互独立。
- 现令 $g(U_1) = U_1 + \mu$ 和 $h(V_1, \cdots, V_n) = (n-1)^{-1} \sum_{r=1}^{n} V_r^2$。则 $g(U_1)$ 和 $h(V_1, \cdots, V_n)$ 相互独立。换言之，S_n^2 和 $\bar{X}_n$ 相互独立，

证毕。

- 还有另一方法可证明 S_n^2 和 $\bar{X}_n$ 相互独立。这是一种富有启发式的证明。

另一种证明 (定理 6.7):

- 令
  - ✓ $\boldsymbol{X} = (X_1, \cdots, X_n)'$ 为 n 维列向量,
  - ✓ $\boldsymbol{l} = (1, \cdots, 1)'$ 为每个元素都是常数 1 的 n 维列向量,
  - ✓ $\boldsymbol{I}$ 为 $n \times n$ 单位方阵,
  - ✓ $\boldsymbol{A}'$ 表示向量或矩阵 $\boldsymbol{A}$ 的转置。
  - ✓ 定义 $n \times n$ 矩阵

$$\boldsymbol{M} = \boldsymbol{I} - \frac{1}{n}\boldsymbol{l}\boldsymbol{l}'$$

证明 (Cont.):

- 注意 $\boldsymbol{M}^2 = \boldsymbol{M}$ 且 $\boldsymbol{M}' = \boldsymbol{M}$。则有

$$\bar{X}_n = \frac{\boldsymbol{l}'\boldsymbol{X}}{n}$$

$$\begin{aligned}(n-1)S_n^2 &= (\boldsymbol{M}\boldsymbol{X})'(\boldsymbol{M}\boldsymbol{X}) \\ &= \boldsymbol{X}'\boldsymbol{M}^2\boldsymbol{X} \\ &= \boldsymbol{X}'\boldsymbol{M}\boldsymbol{X}\end{aligned}$$

证明 (Cont.):

- 为证明 S_n^2 和 $\bar{X}_n$ 相互独立，只需证明随机变量 $\boldsymbol{l}'\boldsymbol{X}$ 和 n 维随机向量 $\boldsymbol{MX}$ 相互独立。
- 令

$$\begin{aligned}\boldsymbol{Z} &= \begin{pmatrix}\boldsymbol{l}'\boldsymbol{X}\\ \boldsymbol{MX}\end{pmatrix}\\ &= \begin{pmatrix}\boldsymbol{l}'\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{X}\\ &= \boldsymbol{AX}\end{aligned}$$

- 其中，$\boldsymbol{A}$ 为 $(n+1)\times n$ 的矩阵。

证明 (Cont.)：

- 因为 $\boldsymbol{Z}$ 是 $\boldsymbol{X}$ 的线性组合，且 $\boldsymbol{X} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$ 是 IID 正态随机向量，故 $\boldsymbol{Z}$ 服从多元正态分布。又因为 $\boldsymbol{l}'\mathbf{M} = \mathbf{0}$ (请证明)，$\boldsymbol{l}'\boldsymbol{X}$ 和 $\boldsymbol{MX}$ 的方差-协方差矩阵

$$\begin{aligned}\operatorname{cov}(\boldsymbol{l}'\boldsymbol{X}, \boldsymbol{MX}) &\equiv E\{[\boldsymbol{l}'\boldsymbol{X} - E(\boldsymbol{l}'\boldsymbol{X})][\boldsymbol{MX} - E(\boldsymbol{MX})]'\} \\ &= E\{\boldsymbol{l}'[\boldsymbol{X} - E(\boldsymbol{X})][\boldsymbol{X} - E(\boldsymbol{X})]'\boldsymbol{M}'\} \\ &= \boldsymbol{l}'E\{[\boldsymbol{X} - E(\boldsymbol{X})][\boldsymbol{X} - E(\boldsymbol{X})]'\}\boldsymbol{M} \\ &= \boldsymbol{l}'\sigma^2\boldsymbol{I}\boldsymbol{M} \\ &= \mathbf{0}\end{aligned}$$

- 因为 $\boldsymbol{l}'\boldsymbol{X}$ 和 $\boldsymbol{MX}$ 服从联合正态分布且二者不相关，故 $\boldsymbol{l}'\boldsymbol{X}$ 和 $\boldsymbol{MX}$ 相互独立。

证毕。

定理 6.8

假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为 IID $N(\mu, \sigma^2)$ 随机样本。则对所有 $n > 1$，

$$\operatorname{var}\left(S_n^2\right) = \frac{2\sigma^4}{n-1}$$

证明：

- 因为

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

- 且 χ_{n-1}^2 的方差为 $2(n-1)$，则有

$$\operatorname{var}\left[\frac{(n-1)S_n^2}{\sigma^2}\right] = 2(n-1)$$

证明 (Cont.)：

或

$$\frac{(n-1)^2}{\sigma^4}\mathrm{var}\left(S_n^2\right)=2(n-1)$$

- 因此 $\mathrm{var}(S_n^2)=2\sigma^4/(n-1)$。

证毕。

- $\mathrm{var}(S_n^2)=2\sigma^4/(n-1)$ 和 $E(S_n^2)=\sigma^2$ 表明，当 $n\to\infty$ 时

$$\begin{aligned}\mathrm{MSE}\left(S_n^2\right)&=E\left(S_n^2-\sigma^2\right)^2\\&=\mathrm{var}\left(S_n^2\right)\\&=\frac{2\sigma^4}{n-1}\to 0\end{aligned}$$

- 因此，当 $n\to\infty$ 时，样本方差 S_n^2 和 σ^2 之间的差距越来越小。换言之，当 n 不断增大时，S_n^2 越来越趋近 σ^2。

目　录

定义 6.6

[学生 t- 分布]：令 $U \sim N(0,1)$，$V \sim \chi_\nu^2$，且 U 和 V 相互独立。则随机变量

$$T = \frac{U}{\sqrt{V/\nu}} \sim \frac{N(0,1)}{\sqrt{\chi_\nu^2/\nu}}$$

服从自由度 ν 的学生 t –分布，记作 $T \sim t_\nu$。

- **学生 t- 分布的 PDF 如下**

$$f_T(t)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}\frac{1}{(\nu\pi)^{1/2}}\frac{1}{(1+t^2/\nu)^{(\nu+1)/2}},-\infty<t<\infty$$

- 可通过先求如下**二元变换**的联合 PDF $f_{TR}(t,r)$

$$\begin{cases}T=U/\sqrt{V/\nu}\\R=U\end{cases}$$

- 再积分消去 R 求得上述结果。

引理 6.3

[学生 t- 分布的性质]:

(1) t_ν 的 PDF 关于 0 对称;

(2) t_ν 分布的尾部比 $N(0,1)$ 更厚 (参见图 6.5);

(3) 只存在前 $\nu-1$ 阶矩。特别地，当 $\nu>2$ 时，均值 $\mu=0$，方差 $\sigma^2=\nu/(\nu-2)$。对任意给定 ν，MGF 不存在;

(4) 当 $\nu=1$ 时，t_1 ~柯西分布 $\text{Cauchy}(0,1)$;

(5) 当 $\nu\to\infty$ 时，$t_\nu\to N(0,1)$。

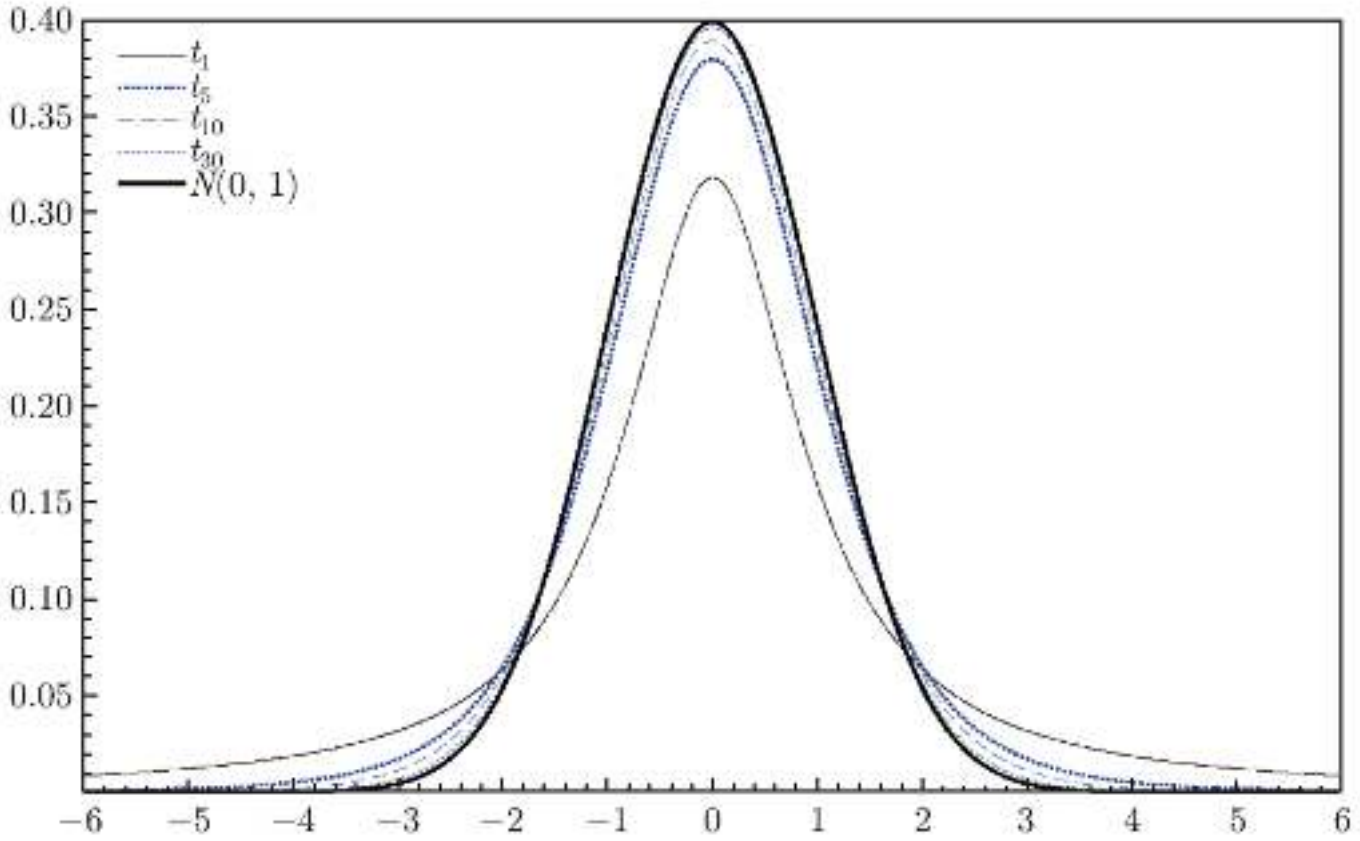

图 6.5：t_1, t_5, t_{10}, t_{30} 以及 $N(0,1)$ 的 PDF

学生 t-分布在统计推断中十分重要

- 当 $\boldsymbol{X}^n$ 为 IID $N(\mu, \sigma^2)$ 随机样本时，对所有 $n \geq 1$ 有

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

- 这是一个非常重要的结果，它可用于当总体方差 σ^2 已知时对未知总体均值 μ 进行**置信区间估计** (confidence interval estimation) 和**假设检验** (hypothesis testing)。
- 然而，在绝大多数实际应用中，一个主要困难是总体标准差 σ 是未知的，因此需要用样本标准差 S_n 代替 σ。这一替代改变了统计量

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

的抽样分布，如以下定理 6.9 所示。

定理 6.9

假设 $\boldsymbol{X}^n = (X, \cdots, X_n)$ 为来自 $N(\mu, \sigma^2)$ 总体的 IID 随机样本。则对所有 $n > 1$，**标准化样本方差** (standardized sample mean)

$$\frac{\bar{X}_n - \mu}{\frac{S_n}{\sqrt{n}}} = \frac{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\frac{(n-1)S_n^2}{\sigma^2}/(n-1)}} \sim \frac{N(0,1)}{\sqrt{\chi^2_{n-1}/(n-1)}} \sim t_{n-1}$$

其中 t_{n-1} 是自由度为 $n-1$ 的学生 t- 分布。

证明：

- 令 $U=(\bar{X}_n-\mu)/(\sigma/\sqrt{n})$ 和 $V=(n-1)S_n^2/\sigma^2$。则

$$U\sim N(0,1),\ V\sim \chi^2_{n-1}$$

- 同时，由定理 6.7 可知 $\bar{X}_n$ 和 S_n^2 相互独立。则

$$\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}=\frac{(\bar{X}_n-\mu)/(\sigma/\sqrt{n})}{\sqrt{(n-1)S_n^2/[\sigma^2(n-1)]}}\sim t_{n-1}$$

证毕。

例 6.7：[关于总体均值 μ 的置信区间估计]

- 假设 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 是来自总体为 $N(\mu, \sigma^2)$ 分布的 IID 随机样本，其中 μ 和 σ^2 均未知。我们的主要目的是在 $(1-\alpha)100\%$ 置信水平上构建 μ 的置信区间估计量。
- 给定 $0 < \alpha < 1$，μ 的 $(1-\alpha)100\%$ **置信区间估计量** (confidence interval estimator) 定义为随机区间 $[\hat{L}, \hat{U}]$，使得真实总体均值 μ 落入区间 $[\hat{L}, \hat{U}]$ 的概率等于 $1-\alpha$，即

$$P(\hat{L} < \mu < \hat{U}) = 1 - \alpha$$

例 6.7 (Cont.)：

- 当 σ^2 未知时，为了构造一个关于 μ 的区间估计量，定义学生 t_{n-1} 分布的**右侧临界值** (upper-tailed critical value) $C_{t_{n-1},\alpha}$ 如下 $P\left(t_{n-1} > C_{t_{n-1},\alpha}\right) = \alpha$

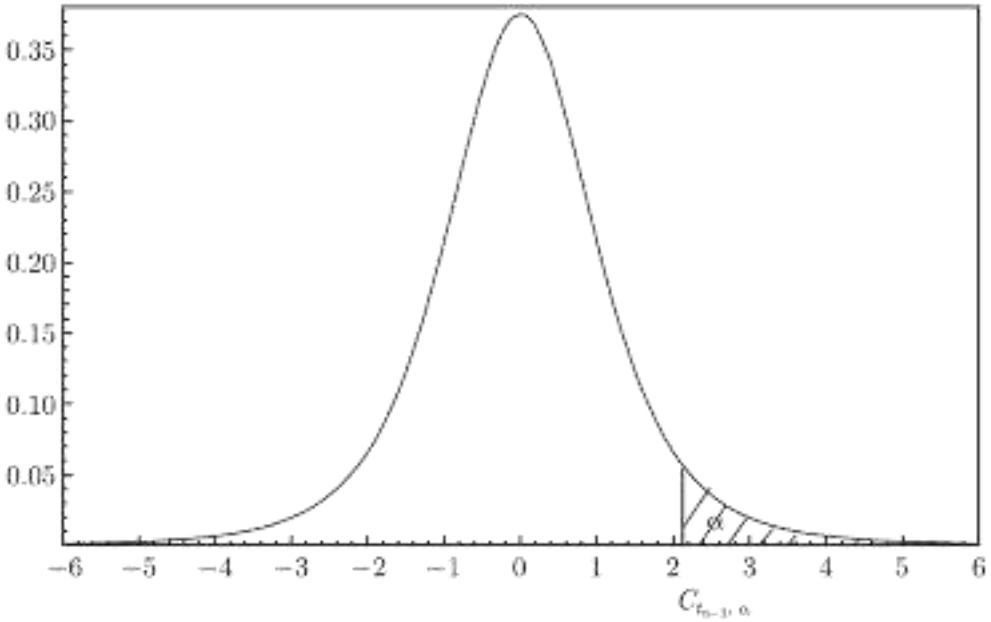

图 6.6：学生 t_{n-1} 分布的右侧临界值 $C_{t_{n-1},\alpha}$

例 6.7 (Cont.)：

- 如图 6.6 所示。根据定理 6.9 以及学生 t-分布的对称性，有

$$P\left[\left|\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n}\right|>C_{t_{n-1},\frac{a}{2}}\right]=\alpha$$

或等价地

$$P\left[\left|\frac{\sqrt{n}(\bar{X}_n-\mu)}{S_n}\right|\leq C_{t_{n-1},\frac{a}{2}}\right]=1-\alpha$$

- 从而，μ 的 $(1-\alpha)100\%$ 置信区间估计量为

$$P\left(\bar{X}_n-\frac{S_n}{\sqrt{n}}C_{t_{n-1},\frac{\alpha}{2}}<\mu<\bar{X}_n+\frac{S_n}{\sqrt{n}}C_{t_{n-1},\frac{\alpha}{2}}\right)=1-\alpha$$

例 6.7 (Cont.)：

- 给定一个随机样本 $\boldsymbol{X}^n$，**随机区间估计量**

$$\left[\bar{X}_n - \frac{S_n}{\sqrt{n}} C_{t_{n-1},\frac{a}{2}}, \bar{X}_n + \frac{S_n}{\sqrt{n}} C_{t_{n-1},\frac{a}{2}}\right]$$

是完全可计算的。

- 此处统计量

$$\frac{\bar{X}_n - \mu}{S_n/\sqrt{n}}$$

的抽样分布在确定临界值 $C_{t_{n-1},\frac{\alpha}{2}}$ 时发挥了至关重要的作用，故在确定置信区间估计量时也十分关键。

例 6.8：[关于总体均值的假设检验：t-检验]

- 假设有一个来自总体为 $N(\mu, \sigma^2)$ 分布的 IID 随机样本 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$，其样本容量为 n。我们的目的是检验如下参数假设

$$\mathbb{H}_0 : \mu = \mu_0$$

  其中 μ_0 为给定 (已知) 常数 (例如 $\mu_0 = 0$)。
- 那么如何检验这个参数假设呢?

例 6.8 (Cont.)：

- 为检验假设 $\mathbb{H}_0$，考察统计量

$$\bar{X}_n - \mu_0 = (\bar{X}_n - \mu) + (\mu - \mu_0)$$

  - ✓ 当 $\mathbb{H}_0$ 为真时，$\mu = \mu_0$。就均方误而言，当 $n \to \infty$ 时，有

$$\bar{X}_n - \mu_0 = \bar{X}_n - \mu \to 0$$

  - ✓ 若 $\mathbb{H}_0$ 为假，即 $\mu \neq \mu_0$。就均方误而言，当 $n \to \infty$ 时，有

$$\bar{X}_n - \mu_0 = (\bar{X}_n - \mu) + (\mu - \mu_0)$$
$$\to \mu - \mu_0 \neq 0$$

- 因此，对 $\mathbb{H}_0$ 的检验可基于统计量 $\bar{X}_n - \mu_0$。
  - ✓ 若 $\bar{X}_n - \mu_0$ 足够小，则 $\mathbb{H}_0$ 为真；
  - ✓ 若 $\bar{X}_n - \mu_0$ 的绝对值足够大，则 $\mathbb{H}_0$ 为假。

◆ **问题 6.4**

$\bar{X}_n - \mu_0$ 和零之间的距离为多大时才可认为其绝对值“足够大”呢?

- 这可由 $\bar{X}_n - \mu_0$ 的抽样分布决定。从 $\bar{X}_n - \mu_0$ 的抽样分布中，可找出称为**临界值** (critical value) 的**门槛值** (threshold value)，并据此通过比较判断 $\bar{X}_n - \mu_0$ 是否足够大。
- 设 $\boldsymbol{X}^n \sim \text{IID } N(\mu, \sigma^2)$。定理 6.2 和 6.3 已证，对每个正整数 n，有

$$\bar{X}_n - \mu \sim N\left(0, \frac{\sigma^2}{n}\right)$$

- 则

$$\bar{X}_n - \mu_0 = (\bar{X}_n - \mu) + (\mu - \mu_0) \sim N\left(\mu - \mu_0, \frac{\sigma^2}{n}\right)$$

问题 6.4 (Cont.)

- 因此，**标准化随机变量**

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} + \frac{\sqrt{n}(\mu - \mu_0)}{\sigma}$$

$$\sim N\left[\frac{\sqrt{n}(\mu - \mu_0)}{\sigma}, 1\right]$$

- 当假设 $\mathbb{H}_0$ 为真时，

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$$

- 这表明当 $\mathbb{H}_0$ 为真时，比值 $(\bar{X}_n - \mu_0)/(\sigma/\sqrt{n})$ 以很大的概率取一个较小的有限值，而 $(\bar{X}_n - \mu_0)/(\sigma/\sqrt{n})$ 取很大值的概率则很小。

问题 6.4 (Cont.)

- 当 $\mathbb{H}_0$ 为假时，则当 $n \to \infty$ 时，有大概率出现

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}} \to \infty$$

- 因此，可通过检查 $(\bar{X}_n - \mu_0)/(\sigma/\sqrt{n})$ 的绝对值是否足够大来判断 $\mathbb{H}_0$ 是否为真。

问题 6.4 (Cont.)

- 但是，比值

$$\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$$

- 并非统计量，因为其中包含未知参数 σ (注意 μ_0 为给定常数值，例如 $\mu_0 = 0$，故 μ_0 不存在问题)，需要以 σ 的估计量代替它，例如样本标准差 S_n。
- 从而有以下的**可行统计量** (feasible statistic)

$$T(\boldsymbol{X}^n) = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$$

- 然而，当用 S_n 替代 σ 后，统计量 $T(\boldsymbol{X}^n)$ 的分布不再是 $N(0,1)$，而变成自由度为 $n-1$ 的学生 t –分布。

问题 6.4 (Cont.)

- 在假设 $\mathbb{H}_0$: $\mu = \mu_0$ 下，对所有 $n > 1$，有

$$T(\boldsymbol{X}^n) \sim t_{n-1}$$

✓ 这是因为在原假设 $\mathbb{H}_0$ 成立时

$$\begin{aligned} T(\boldsymbol{X}^n) &= \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}} \\ &= \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} + \frac{\sqrt{n}(\mu - \mu_0)}{S_n} \\ &= \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \\ &\sim t_{n-1} \end{aligned}$$

✓ 因此，当 $\mathbb{H}_0$ 为真时，t-检验统计量 $T(\boldsymbol{X}^n)$ 以很大的概率取较小的有限值，而取很大值的概率则很小。

问题 6.4 (Cont.)

- 当 $\mathbb{H}_0$：$\mu = \mu_0$ 为假时，即当 $\mu \neq \mu_0$，$n \to \infty$ 时，有

$$T(\boldsymbol{X}^n) = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} + \frac{\sqrt{n}(\mu - \mu_0)}{S_n} \to \infty$$

  - ✓ 换言之，在 $\mathbb{H}_0$ 的备择假设下，当 $n \to \infty$ 时，统计量 $T(\boldsymbol{X}^n)$ 以接近 1 的概率发散到无穷。

基于临界值的 t-检验决策准则：

- (1) 在预设显著水平 $\alpha \in (0,1)$ 下，若 t-检验统计量的绝对值

$$|T(\boldsymbol{X}^n)| > C_{t_{n-1},\frac{\alpha}{2}}$$

- 则拒绝原假设 $\mathbb{H}_0: \mu = \mu_0$，其中 $C_{t_{n-1},\frac{\alpha}{2}}$ 是当显著水平为 $\frac{\alpha}{2}$ 时学生 t_{n-1} 分布的右侧临界值，由 $P\left(t_{n-1} > C_{t_{n-1},\frac{\alpha}{2}}\right) = \frac{\alpha}{2}$ 决定。

基于临界值的 t- 检验决策准则 (Cont.):

- (2) 在显著水平 α 上，若 $|T(\boldsymbol{X}^n)| \leq C_{t_{n-1},\frac{\alpha}{2}}$，则无法拒绝原假设 $\mathbb{H}_0$ 。
- 基于临界值的 t-检验决策准则对应的**拒绝域**和**接受域**见下图:
- 直观上，t-检验决策准则表示，
  - ✓ 当 $|T(\boldsymbol{X}^n)| > C_{t_{n-1},\frac{\alpha}{2}}$ 时，说明 $\bar{X}^n - \mu_0$ 显著不为 0，因此拒绝 $\mathbb{H}_0$。
  - ✓ 若 $|T(\boldsymbol{X}^n)| \leq C_{t_{n-1},\frac{\alpha}{2}}$ ，则 $\bar{X}^n - \mu_0$ 并不显著异于 0，故无法拒绝 $\mathbb{H}_0$ 。

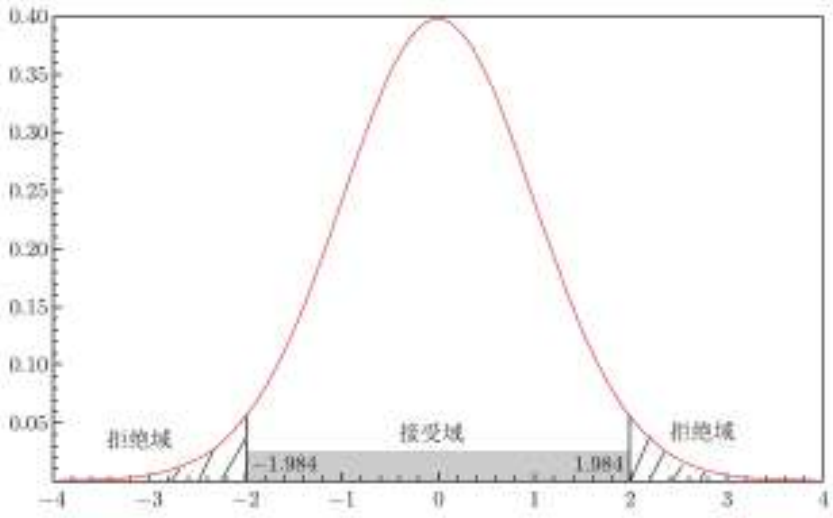

图 6.7：$n = 30, \alpha = 5\%$ 时 t-检验的拒绝域与接受域

第一类错误 (Type I error) & 第二类错误 (Type II error)

- 在使用从样本容量为 n 的随机样本 $\boldsymbol{X}^n$ 所生成的数据检验 $\mathbb{H}_0$ 时，存在**两类错误**。
  - ✓ **第一类错误**：$\mathbb{H}_0$ 为真，但被错误拒绝。
    - ➢ 发生原因：检验统计量 $T(\boldsymbol{X}^n)$ 在 $\mathbb{H}_0$ 假设下服从学生 t_{n-1} 分布，而该分布有一个无界的支撑。因此，存在一个小概率使得 $T(\boldsymbol{X}^n)$ 在原假设 $\mathbb{H}_0$下仍可能取大于临界值的值。
  - ✓ **第二类错误**：$\mathbb{H}_0$ 为假，但被错误接受。

第一类错误 & 第二类错误 (Cont.):

- 显著水平 α 控制**第一类错误**的概率。常用的显著水平为 10%，5% 或 1%。若

$$P\left[|T(\boldsymbol{X}^n)| > C_{t_{n-1},\frac{a}{2}} \middle| \mathbb{H}_0\right] = \alpha$$

则该检验决策准则称为**尺度** (size) α 的检验或 α 尺度的检验。

- 另一方面，概率

$$P\left[|T(\boldsymbol{X}^n)| > C_{t_{n-1},\frac{a}{2}} \middle| \mathbb{H}_0\text{为假}\right]$$

称为尺度为 α 的 t-检验的**功效函数** (power function)。

- **第二类错误**的概率：

$$P\left[|T(\boldsymbol{X}^n)| \leq C_{t_{n-1},\frac{a}{2}} \middle| \mathbb{H}_0\text{为假}\right]$$

第一类错误 & 第二类错误 (Cont.):

- 当 n 有限时，由于随机样本 $\boldsymbol{X}^n$ 提供的信息有限，因此犯第一类错误和第二类错误均是无法避免的，两者之间通常存在此消彼长的关系（问题：为什么？）。
- 在实际应用中，一般预设第一类错误的水平，并尽量使第二类错误的概率最小。

- 当 n 很大时，标准正态分布 $N(0,1)$ 与 t-分布的临界值很接近，因此可用 $N(0,1)$ 的临界值近似 t-分布的临界值。
- 假设 $\boldsymbol{X}^n$ 为 IID $N(\mu,\sigma^2)$ 随机样本，则当 $n\to\infty$

$$\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}\sim t_{n-1}\to N(0,1)$$

- 实际应用中，若 $n-1\geq 30$，正态分布可很好地近似 t_{n-1}。

使用 P- 值作为 t- 检验的决策准则

- 给定任意观测数据集 $\boldsymbol{x}^n$，t-检验统计量 $T(\boldsymbol{X}^n)$ 有一个对应的实现值。

$$T(\boldsymbol{x}^n) = \frac{\bar{x}_n - \mu_0}{s_n/\sqrt{n}}$$

- 则概率

$$p(\boldsymbol{x}^n) = P[|T(\boldsymbol{X}^n)| > |T(\boldsymbol{x}^n)| \mid \mathbb{H}_0] = P[|t_{n-1}| > |T(\boldsymbol{x}^n)|]$$

称为当给定观测数据集 $\boldsymbol{x}^n$ 时，t-检验统计量 $T(\boldsymbol{X}^n)$ 的 P-值。

使用 P-值作为 t-检验的决策准则 (Cont.):

- **P-值的含义：** P- 值可视为当 $\mathbb{H}_0$ 为真时，t- 检验统计量 $T(\boldsymbol{X}^n)$ 大于观测值 $T(\boldsymbol{x}^n)$ 的概率。当观测值 $T(\boldsymbol{x}^n)$ 较大时，$p(\boldsymbol{x}^n)$ 将较小。
- 上述基于临界值的决策准则等价于以下基于 **P- 值的决策准则**：
  - ✓ (1) 在显著水平 α 上，若 $p(\boldsymbol{x}^n) < \alpha$，则**拒绝原假设** $\mathbb{H}_0$；
  - ✓ (2) 在显著水平 α 上，若 $p(\boldsymbol{x}^n) \geq \alpha$，则**无法拒绝原假设** $\mathbb{H}_0$。
- P- 值包含的信息比基于临界值的 t- 检验决策准则更多。

统计显著性 vs. 经济显著性

- **统计显著性 (statistically significant)**：从统计角度看，任意对 $\mathbb{H}_0$ 的偏离 (即任意 $\mu - \mu_0$ 之差)，不论多小，当样本容量 n 足够大时都会拒绝 $\mathbb{H}_0$。
- **经济显著性 (economically significant)**：$\mu - \mu_0$ 一个较小的偏差从经济角度看可能并不具有重要的实际意义。
  - ✓ 例如，投资者可能关心某共同基金的预期收益率 (μ) 是否与一个预设的回报率 (μ_0) 有显著不同，$\mu - \mu_0$ 之差需足够大才会考虑投资该共同基金，因为存在交易成本。

统计显著性 vs. 经济显著性 (Cont.)：

- **一个经济意义上并不显著的差别效应可能在统计意义上显著。**
  - ✓ 如果样本容量 n 足够大，像 t-检验这样的统计检验会在样本容量 n 足够大时拒绝一个很小的 $\mu - \mu_0$ 之差。
- **一个经济意义上重要的效应差别可能在统计意义上并不显著。**
  - ✓ 当样本容量 n 很小时，这可能会发生，从而导致有很大的概率发生第二类错误。
- **没有统计显著性也可能是因为统计假设不能很好地反映重要的经济效应。**比如，线性统计模型的系数可能无法有效刻画非线性效应。在这种情况下，即使样本容量 n 很大，系数也可能不具有统计显著性。

目　录

定理 6.7

[$\mathcal{F}$ -分布]：令 U 和 V 是自由度分别为 p 和 q 的两个独立卡方随机变量，则

$$F = \frac{U/p}{V/q} \sim \mathcal{F}_{p,q}$$

服从自由度为 p 和 q 的 $\mathcal{F}$-分布。

问题 6.5

$\mathcal{F}_{p,q}$-分布的 PDF 是什么呢?

- $\mathcal{F}_{p,q}$-分布的 PDF 为

$$f_F(x)=\frac{\Gamma\left(\frac{p+q}{2}\right)}{\Gamma\left(\frac{p}{2}\right)\Gamma\left(\frac{q}{2}\right)}\left(\frac{p}{q}\right)^{p/2}\frac{x^{(p/2)-1}}{[1+(p/q)x]^{(p+q)/2}},0<x<\infty$$

- 上述 PDF 可通过以下的二元变换

$$\begin{cases}F=(U/p)/(V/q)\\G=U\end{cases}$$

- 先求得 (F,G) 的联合 PDF，然后积分消去 G。

不同自由度 (p,q) 下的 $\mathcal{F}$-分布

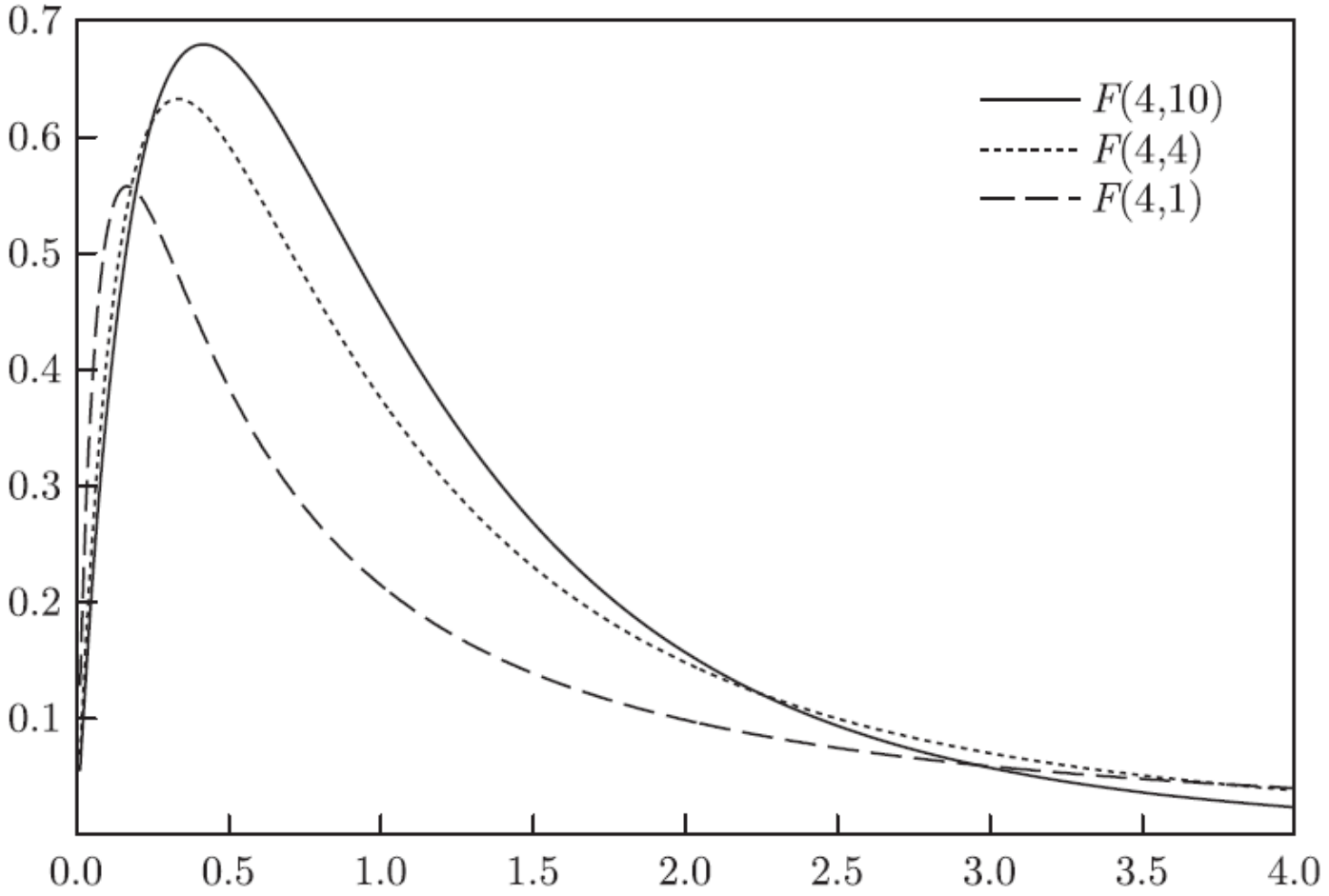

图 6.8：不同自由度 (p,q) 下的 $\mathcal{F}$-分布

引理 6.4

[$\mathcal{F}_{p,q}$– 分布的性质]:

(1) 若 $X \sim \mathcal{F}_{p,q}$，则 $X^{-1} \sim \mathcal{F}_{q,p}$;

(2) 若 $X \sim t_q$，则 $X^2 \sim \mathcal{F}_{1,q}$;

(3) 若 $q \to \infty$，则 $p\mathcal{F}_{p,q} \to \chi_p^2$。

证明:

- 结果(1) 可从 F 随机变量的定义得到。对结果(2)，随机变量 t_q 定义为

$$t_q \sim \frac{Z}{\sqrt{\chi_q^2/q}}$$

证明 (Cont.):

- 其中 $Z \sim N(0,1)$，且与 χ_q^2 相互独立。因此有

$$t_q^2 \sim \frac{\chi_1^2/1}{\chi_q^2/q} \sim \mathcal{F}_{1,q}$$

- **证毕。**

例 6.9：[总体方差相等的假设检验 (Hypothesis Testing on Equality of Population Variances)]

- 令
  - ✓ $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自总体为正态分布 $N(\mu_X, \sigma^2{}_X)$，样本容量为 n 的 IID 随机样本，
  - ✓ $\boldsymbol{Y}^{\boldsymbol{m}} = (Y_1, \cdots, Y_m)$ 为来自总体为 $N(\mu_Y, \sigma_Y^2)$ 分布，样本容量为 m 的 IID 随机样本。
- 若我们对比较总体的变异性感兴趣，即检验原假设 $\mathbb{H}_0: \sigma_X^2 = \sigma_Y^2$ 是否成立，则可考虑基于如下的样本方差比的检验统计量

$$\frac{S_X^2}{S_Y^2}$$

例 6.9 (Cont.):

- 在均方误意义上，当 $n \to \infty$ 时有 $S_X^2 \to \sigma_X^2$，当 $m \to \infty$ 时有 $S_Y^2 \to \sigma_Y^2$。因此当 $n,\ m \to \infty$ 时

$$\frac{S_X^2}{S_Y^2} \to \frac{\sigma_X^2}{\sigma_Y^2}$$

- 在 $\mathbb{H}_0 : \sigma_X^2 = \sigma_Y^2$ 的假设下，有

$$\begin{aligned} \frac{S_X^2}{S_Y^2} &= \frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2} \\ &= \frac{\dfrac{(n-1)\,S_X^2/\sigma_X^2}{n-1}}{\dfrac{(m-1)\,S_Y^2/\sigma_Y^2}{m-1}} \end{aligned} \qquad \begin{aligned} &\sim \frac{\chi_{n-1}^2/(n-1)}{\chi_{m-1}^2/(m-1)} \\ &\sim \mathcal{F}_{n-1,m-1} \end{aligned}$$

例 6.9 (Cont.):

- 若 $\mathbb{H}_0$ 为假，即 $\sigma_X^2 \neq \sigma_Y^2$，则 $\frac{S_X^2}{S_Y^2} \neq \frac{S_X^2/\sigma_Y^2}{S_Y^2/\sigma_Y^2} \sim \mathcal{F}_{n-1,m-1}$。
- 因此，通过检验 S_X^2/S_Y^2 是否服从 $\mathcal{F}_{n-1,m-1}$，可判断方差是否相等。
- 因为 $\mathcal{F}$-分布和样本方差紧密相关，故也常称为**方差比分布** (variance ratio distribution)。
- 一个服从 $\mathcal{F}$-分布的随机变量不一定是随机样本的方差比。

例 6.9 (Cont.):

- F- 检验是经典数理统计学和经典计量经济学的一个重要检验，其中 S_X^2 和 S_Y^2 可分别进一步推广为:
  - ✓ 受约束回归模型 (restricted regression model) 的残差平方和
  - ✓ 无约束回归模型 (unrestricted regression model) 的残差平方和

例 6.9 (Cont.):

- 例如，考察经典线性回归模型

$$Y_i = X_i'\beta + Z_i'\gamma + \varepsilon_i$$

- 其中
  - ✓ β 为 $p \times 1$ 维参数向量，
  - ✓ γ 为 $q \times 1$ 维参数向量，
  - ✓ $\{\varepsilon_i\}$ 为 IID $N(0, \sigma_\varepsilon^2)$ 随机变量序列，且与 $(X_1, \cdots, X_n)$ 和 $(Z_1, \cdots, Z_n)$ 相互独立。

例 6.9 (Cont.)：

- 假设目的是检验原假设 $\mathbb{H}_0$: $\gamma = \mathbf{0}$ 是否成立。
  - ✓ 在 $\mathbb{H}_0$ 假设下，Z_i 对条件均值 $E(Y_i \mid X_i, Z_i)$ 没有影响。
  - ✓ 在 $\mathbb{H}_0$ 的备择假设下，Z_i 对条件均值 $E(Y_i \mid X_i, Z_i)$ 有影响，因此，有约束线性回归模型 $Y_i = X_i'\beta + \varepsilon_i$ 存在遗漏变量问题。

例 6.9 (Cont.):

- 为了检验原假设 $\mathbb{H}_0$ ， 可进行两次最小二乘法回归。

(1) 无约束回归模型

$$Y_i = X_i'\beta + Z_i'\gamma + \varepsilon_i$$

- 其**最小二乘法估计量**为

$$(\hat{\beta}, \hat{\gamma}) = \arg \min_{\beta \in \mathbb{R}^p, \gamma \in \mathbb{R}^q} \sum_{i=1}^{n} (Y_i - X_i'\beta - Z_i'\gamma)^2$$

- 相应的**残差方差估计量**为

$$S_U^2 = \frac{1}{n-p-q} \sum_{i=1}^{n} \left(Y_i - X_i'\hat{\beta} - Z_i'\hat{\gamma}\right)^2$$

- 其中下标 U 表示无约束回归模型。

例 6.9 (Cont.):

(2) 有约束回归模型

$$Y_i = X_i'\beta + \mu_i$$

- 其**最小二乘法估计量**为

$$\tilde{\beta} = \arg\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (Y_i - X_i'\beta)^2$$

- 相对应的**残差方差估计量**为

$$S_R^2 = \frac{1}{n-p} \sum_{i=1}^{n} \left(Y_i - X_i'\tilde{\beta}\right)^2$$

- 其中，下标 R 表示有约束回归模型。

例 6.9 (Cont.)：

- 为检验 $\mathbb{H}_0$，可比较残差方差估计量 S_U^2 和 S_R^2。
  - ✓ 在 $\mathbb{H}_0$ 假设下，二者将收敛于同一极限。
  - ✓ 在备择假设下，有约束回归模型因遗漏了变量 Z_i 对 Y_i 的影响，从而引起回归模型误设，导致 S_U^2 和 S_R^2 分别收敛于不同极限，且 $\lim_{n\to\infty}\left(S_R^2/S_U^2\right) > 1$。
- 因此关于 $\mathbb{H}_0$ 的检验统计量可构造如下：

$$F = \frac{\left[\left(n-p\right)S_R^2 - \left(n-p-q\right)S_U^2\right]/q}{\left(n-p-q\right)S_U^2/\left(n-p-q\right)}$$

例 6.9(Cont.):

- 此 F 统计量非负，因为无约束回归模型的残差平方和总小于或等于有约束回归模型的残差平方和。
- 可以证明，当 $\mathbb{H}_0$ 为真时，在适当的正则条件下

$$F \sim \frac{\chi_q^2/q}{\chi_{n-p-q}^2/(n-p-q)} \sim \mathcal{F}_{q,n-p-q}$$

- 因为在备择假设下 $\lim_{n\to\infty}\left(S_R^2/S_U^2\right) > 1$，因此需要使用 $\mathcal{F}_{p,q}$-分布的右侧临界值。
- 更多讨论参见第十章。

目　录

6.6 充分统计量

- **“KISS”原则：** Keep It Sophistically Simple，即尽量用最简单的模型刻画数据所包含的重要信息。

◆ **问题**

假设要使用来自总体 $f_X(x) = f(x,\theta)$ 的随机样本 $\boldsymbol{X}^n$ 所生成的数据集对参数 θ 进行推断。那么，在什么条件下随机样本 $\boldsymbol{X}^n$ 中关于 θ 的信息可用 $\boldsymbol{X}^n$ 的某个低维 (low dimensional) 函数完全概括，如统计量 $T(\boldsymbol{X}^n)$?

- 假设一个随机试验产生了随机样本 $\boldsymbol{X}^n$ 的一个实现值 $\boldsymbol{x}^n$，同时假设：
  - ✓ 某统计学家 A 观察到数据集 $\boldsymbol{x}^n$，
  - ✓ 而另一统计学家 B 仅观察到统计量 $t = T(\boldsymbol{X}^n)$ 的值。
- 一般而言，A 比 B 拥有更多关于未知参数 θ 值的信息。
- 但是，有可能存在 B 和 A 实际上拥有关于 θ 的同样多样本信息的情形。
  - ✓ 若统计量 $T(\boldsymbol{X}^n)$ 概括了随机样本 $\boldsymbol{X}^n$ 中的关于 θ 的所有信息时，那么数据集 $\boldsymbol{x}^n$ 的个体值便没有提供关于 θ 的更多信息。

- 具有这种理想性质的统计量 $T(\boldsymbol{X}^n)$ 称为 θ 的**充分统计量** (Sufficient Statistic)。
- 充分统计量对参数 θ 的一个重要含义在于只需关注该低维统计量，其最大方便之处是降维，因为原始随机样本 $\boldsymbol{X}^n$ 的维数很高，等于样本容量 n。

定义 6.8

[充分统计量 (Sufficient Statistic)]：令 $\boldsymbol{X}^n$ 为来自以 θ 为参数的某个总体分布的随机样本。给定统计量 $T(\boldsymbol{X}^n)$ 的值，即当 $T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)$ 时，若随机样本 $\boldsymbol{X}^n=\boldsymbol{x}^n$ 的条件概率分布不依赖于 θ 值，即对所有可能的 θ 值，有

$$f_{\boldsymbol{X}^n|T(\boldsymbol{X}^n)}[\boldsymbol{x}^n \mid T(\boldsymbol{x}^n),\theta]=h(\boldsymbol{x}^n)$$

则称统计量 $T(\boldsymbol{X}^n)$ 为 θ 的**充分统计量**。其中，

- 等式左边为给定 $T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)$ 时，$\boldsymbol{X}^n=\boldsymbol{x}^n$ 的条件 PMF/PDF，一般来说依赖于 θ。
- 等式右边 $h(\boldsymbol{x}^n)$ 不依赖于 θ，它只是样本点 $\boldsymbol{x}^n$ 的函数。

充分统计量的含义

- 给定 $T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)$，若随机样本 $\boldsymbol{X}^n=\boldsymbol{x}^n$ 的条件概率 $f_{\boldsymbol{X}^n|T(\boldsymbol{X}^n)}[\boldsymbol{x}^n \mid T(\boldsymbol{x}^n),\theta]$ 不依赖于 θ，则给定 $T(\boldsymbol{x}^n)$ 值，所有使 $T(\boldsymbol{x}^n)=t$ 成立的样本点 $\boldsymbol{x}^n$ 对任意 θ 值均具有相同的概率。
- 换言之，给定 $T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)$ 时，$\boldsymbol{X}^n=\boldsymbol{x}^n$ 的条件分布不依赖于 θ，因此满足数据集 $\boldsymbol{x}^n$ 并没有比统计量 $T(\boldsymbol{x}^n)=t$ 提供更多的关于 θ 的信息。
- 因此，除了 $T(\boldsymbol{x}^n)=t$ 值之外的数据集 $\boldsymbol{x}^n$ 的信息无助于对 θ 的推断。θ 的充分统计量 $T(\boldsymbol{X}^n)$ 已完全捕捉了随机样本 $\boldsymbol{X}^n$ 中与 θ 相关的所有信息。所有从随机样本 $\boldsymbol{X}^n$ 的数据集 $\boldsymbol{x}^n$ 获得的关于 θ 的信息都可从统计量 $T(\boldsymbol{x}^n)$ 获得。

离散情形下的充分统计量 $T(\boldsymbol{X}^n)$

- 首先，充分性意味着对所有 θ 值，随机样本 $\boldsymbol{X}^n$ 基于 $T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)$ 下的条件 PMF

$$f_{\boldsymbol{X}^n|T(\boldsymbol{X}^n)}[\boldsymbol{x}^n \mid T(\boldsymbol{x}^n), \theta] \equiv P_\theta[\boldsymbol{X}^n = \boldsymbol{x}^n \mid T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)] = h(\boldsymbol{x}^n)$$

- 其中，$P_\theta(\cdot)$ 是 $\boldsymbol{X}^n$ 概率分布下的概率测度，通常依赖于参数 θ。

离散情形下的充分统计量 $T(\boldsymbol{X}^n)$ (Cont.)

- 另一方面，一个随机样本 $\boldsymbol{X}^n$ 的全部信息可由 $\boldsymbol{X}^n = \boldsymbol{x}^n$ 的联合概率描述，记作 $P(\boldsymbol{X}^n = \boldsymbol{x}^n) = f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)$。一般而言，该联合概率依赖于 θ。例如，当 $\boldsymbol{X}^n$ 为来自总体 PMF 为 $f(x, \theta)$ 的 IID 随机样本时，有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$$

- 因为 $T(\cdot)$ 是一个函数，从 $\boldsymbol{X}^n = \boldsymbol{x}^n$ 可推出 $T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)$，反之则不成立。
- 因此有 $A = \{\boldsymbol{X}^n = \boldsymbol{x}^n\} \subseteq B = \{T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)\}$，从而 $A = A \cap B$。

离散情形下的充分统计量 $T(X^n)$ (Cont.)

- 由充分性，随机样本 $\boldsymbol{X}^n$ 的联合 PMF

$$
\begin{aligned}
f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) &= P(\boldsymbol{X}^n = \boldsymbol{x}^n) \\
&= P(A \cap B) \\
&= P(A \mid B)P(B) \\
&= P[\boldsymbol{X}^n = \boldsymbol{x}^n \mid T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)]P[T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)] \\
&= h(\boldsymbol{x}^n) f_{T(\boldsymbol{X}^n)}[T(\boldsymbol{x}^n), \theta]
\end{aligned}
$$

- 其中 $f_{T(\boldsymbol{X}^n)}[T(\boldsymbol{x}^n), \theta] \equiv P[T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)]$ 依赖于 θ，而 $h(\boldsymbol{x}^n)$ 不依赖于 θ。

离散情形下的充分统计量 $T(\boldsymbol{X}^n)$ （Cont.）

- 充分统计量 $T(\boldsymbol{X}^n)$ 的抽样分布 $P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]$ 与 θ 有关，而其他部分 $h(\boldsymbol{x}^n)$ 与 θ 无关。
- 因此，若要对 θ 进行推断，只需保留 $T(\boldsymbol{X}^n)$ 的信息，而随机样本 $\boldsymbol{X}^n$ 中其他的信息对于推断 θ 则是多余的。
- 换言之，在推断参数 θ 时，用低维的 $T(\boldsymbol{x}^n)$ 的信息与用高维数据集 $\boldsymbol{x}^n$ 的信息的效果是相同的。

- 例如，第八章将要介绍的极大似然估计方法 (maximum likelihood estimation, MLE) 就是选择 θ 值最大化目标函数——**对数似然函数**

$$\ln f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) = \ln h(\boldsymbol{x}^n) + \ln f_{T(\boldsymbol{X}^n)}[T(\boldsymbol{x}^n), \theta]$$

- 因为上式中第一部分与 θ 无关，故有

$$\hat{\theta}(\boldsymbol{x}^n) \equiv \arg\max_{\theta \in \Theta} \ln f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) = \arg\max_{\theta \in \Theta} \ln f_{T(\boldsymbol{X}^n)}[T(\boldsymbol{x}^n), \theta]$$

其中 Θ 为参数空间。

- 换言之，对 θ 的 MLE 只需最大化充分统计量的对数似然函数 $\ln f_{T(\boldsymbol{X}^n)}[T(\boldsymbol{x}^n), \theta]$。

◆ 问题

如何判断 $T(\boldsymbol{X}^n)$ 是参数 θ 的充分统计量?

定理 6.10

[因子分解定理 (Factorization Theorem)]：令 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)$ 为随机样本 $\boldsymbol{X}^n$ 的联合 PMF/PDF。当且仅当存在函数 $g(t,\theta)$ 和 $h(\boldsymbol{x}^n)$，满足对 $\boldsymbol{X}^n$ 的样本空间中的任意样本点 $\boldsymbol{x}^n$ 以及任意参数值 $\theta \in \Theta$，都有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta) = g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)$$

则统计量 $T(\boldsymbol{X}^n)$ 为 θ 的充分统计量，其中 $g(t,\theta)$ 依赖于参数 θ，但 $h(\boldsymbol{x}^n)$ 不依赖于参数 θ。

证明：

- 此处仅证明离散情形，其中 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)=P(\boldsymbol{X}^n=\boldsymbol{x}^n)$。

(1) [必要性] 当 $T(\boldsymbol{X}^n)$ 为充分统计量时，

- 因为 $\{\boldsymbol{X}^n=\boldsymbol{x}^n\}\subseteq\{T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)\}$，有

$$\{\boldsymbol{X}^n=\boldsymbol{x}^n\}=\{\boldsymbol{X}^n=\boldsymbol{x}^n\}\cap\{T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)\}$$

- 因此

$$\begin{aligned}f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)&=P(\boldsymbol{X}^n=\boldsymbol{x}^n)\\&=P[\boldsymbol{X}^n=\boldsymbol{x}^n,T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]\\&=P[\boldsymbol{X}^n=\boldsymbol{x}^n\mid T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]\\&=h(\boldsymbol{x}^n)P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]\\&=h(\boldsymbol{x}^n)g[T(\boldsymbol{x}^n),\theta]\end{aligned}$$

- 其中 $g[T(\boldsymbol{x}^n),\theta]=P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]$ 和 $h(\boldsymbol{x}^n)=P[\boldsymbol{X}^n=\boldsymbol{x}^n\mid T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]$，后者不依赖于 θ。

证明（Cont.）：

(2) [充分性] 假设有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)=g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)$$

- 将证明条件概率 $P[\boldsymbol{X}^n=\boldsymbol{x}^n|\,T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]$ 不依赖于 θ。
- 因为

$$\{\boldsymbol{X}^n=\boldsymbol{x}^n\}=\{\boldsymbol{X}^n=\boldsymbol{x}^n\}\cap\{T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)\}$$

- 有

$$P[\boldsymbol{X}^n=\boldsymbol{x}^n\mid T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]=\frac{P[\boldsymbol{X}^n=\boldsymbol{x}^n,T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]}{P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]}$$

$$=\frac{P(\boldsymbol{X}^n=\boldsymbol{x}^n)}{P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]}$$

$$=\frac{g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)}{P[T(\boldsymbol{X}^n)=T(\boldsymbol{x}^n)]}$$

证明（Cont.）：

- 现在考察分母

$$
\begin{aligned}
P[T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)] &= \sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} f_{\boldsymbol{X}^n}(\boldsymbol{y}^n, \theta) \\
&= \sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} g[T(\boldsymbol{y}^n), \theta] h(\boldsymbol{y}^n) \\
&= \sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} g[T(\boldsymbol{x}^n), \theta] h(\boldsymbol{y}^n) \\
&= g[T(\boldsymbol{x}^n), \theta] \sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} h(\boldsymbol{y}^n)
\end{aligned}
$$

- 其中求和是针对 $\boldsymbol{X}^n$ 的样本空间（即支撑）中满足约束条件 $T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)$ 的所有可能的样本点 $\{\boldsymbol{y}^n\}$。

证明（Cont.）：

- 则条件概率为

$$P[\boldsymbol{X}^n = \boldsymbol{x}^n \mid T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)] = \frac{g[T(\boldsymbol{x}^n), \theta]h(\boldsymbol{x}^n)}{P[T(\boldsymbol{X}^n) = T(\boldsymbol{x}^n)]}$$

$$= \frac{g[T(\boldsymbol{x}^n), \theta]h(\boldsymbol{x}^n)}{g[T(\boldsymbol{x}^n), \theta]\sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} h(\boldsymbol{y}^n)}$$

$$= \frac{h(\boldsymbol{x}^n)}{\sum_{\{\boldsymbol{y}^n : T(\boldsymbol{y}^n) = T(\boldsymbol{x}^n)\}} h(\boldsymbol{y}^n)}$$

其不依赖于 θ。

证毕。

例 6.10：

- 假设 $\boldsymbol{X}^n \sim \text{IID Bernoulli}(\theta)$，其中 $0 < \theta < 1$。
- 证明 $T(\boldsymbol{X}^n) = n^{-1}\sum_{i=1}^{n} X_i$ 是 θ 的充分统计量。注意 $\theta = E(X_i)$。

解：

- 伯努利随机变量 X_i 的 PMF 为

$$f(x_i, \theta) = \theta^{x_i}(1-\theta)^{1-x_i}$$

其中 x_i 可取值 1 或 0。

解 (Cont.):

- 假设 $\boldsymbol{x}^n$ 为随机样本 $\boldsymbol{X}^n$ 的一个实现值 (即一个数据集)。则有

$$\begin{aligned} P(\boldsymbol{X}^n = \boldsymbol{x}^n) &= \prod_{i=1}^n f(x_i, \theta) \\ &= \prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i} \\ &= \theta^{nT(\boldsymbol{x}^n)}(1-\theta)^{n-nT(\boldsymbol{x}^n)} \\ &= g[T(\boldsymbol{x}^n), \theta]h(\boldsymbol{x}^n) \end{aligned}$$

- 其中 $T(\boldsymbol{X}^n) = n^{-1}\sum_{i=1}^n X_i$, $h(\boldsymbol{x}^n) = 1$, 且 $g[T(\boldsymbol{x}^n), \theta] = \theta^{nT(\boldsymbol{x}^n)}(1-\theta)^{n-nT(\boldsymbol{x}^n)}$。
- 注意，$nT(\boldsymbol{X}^n) = \sum_{i=1}^n X_i$ 也是 θ 的充分统计量。

例 6.11：

- 令 $\boldsymbol{X}^n \sim \text{IID } N(\mu, \sigma^2)$，其中 σ^2 已知。
- 证明 $T(\boldsymbol{X}^n) = \bar{X}_n$, 为 μ 的充分统计量。

解：

- 本例中未知参数 $\theta = \mu$。因为 σ^2 已知，故不再是参数。$\boldsymbol{X}^n$ 的联合 PDF 为

$$
\begin{aligned}
f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \mu) &= \prod_{i=1}^{n} f_{X_i}(x_i, \theta) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{i=1}^{n}(x_i-\bar{x}_n+\bar{x}_n-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{n=1}^{n}(x_i-\bar{x}_n)^2+n(\bar{x}_n-\mu)^2}{2\sigma^2}} \\
&= \left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{n=1}^{n}(x_i-\bar{x}_n)^2}{2\sigma^2}}\right] e^{-\frac{n(\bar{x}_n-\mu)^2}{2\sigma^2}} \\
&= h(\boldsymbol{x}^n) g(\bar{x}_n, \mu)
\end{aligned}
$$

解 (Cont.):

- 其中

$$h(\boldsymbol{x}^n) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{\sum_{n=1}^{n}(x_i-\bar{x}_n)^2}{2\sigma^2}}$$

$$g[T(\boldsymbol{x}^n), \theta] = e^{-\frac{n(\bar{x}_n-\mu)^2}{2\sigma^2}}$$

- 则 $T(\boldsymbol{X}^n)=\bar{X}_n$ 是 μ 的充分统计量。

例 6.12：

- 假设 $\boldsymbol{X}^n \sim \text{IID}\ N(\mu, \sigma^2)$，其中 μ 和 σ^2 均为未知参数。则 $T(\boldsymbol{X}^n) = (\bar{X}_n, S_n^2)$ 为 (μ, σ^2) 的充分统计量。

解：

- 本例中，未知参数 $\theta = (\mu, \sigma^2)$ 是二维向量。

解 (Cont.):

- 随机样本 $\boldsymbol{X}^n$ 的联合 PDF 为

$$
\begin{aligned}
f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{(n-1)[(n-1)^{-1}\sum_{n=1}^{n}(x_i-\bar{x}_n)^2]}{2\sigma^2}-\frac{n(\bar{x}_n-\mu)^2}{2\sigma^2}} \\
&= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{(n-1)S_n^2+n(\bar{x}_n-\mu)^2}{2\sigma^2}} \\
&= g[T(\boldsymbol{x}^n),\theta]h(\boldsymbol{x}^n)
\end{aligned}
$$

  其中对所有 $\boldsymbol{x}^n$，$h(\boldsymbol{x}^n)=1$。

- 则二维统计量 $T(\boldsymbol{X}^n)=(\bar{X}_n,\ S_n^2)$ 是 $\theta=(\mu,\sigma^2)$ 的充分统计量。

- 这个结果解释了为什么经典统计抽样理论主要考虑样本均值 $\bar{X}_n$ 和样本方差 S_n^2。
  - ✓ 因为经典统计抽样理论假设随机样本 $\boldsymbol{X}^n$ 来自正态分布。在此假设条件下，$(\bar{X}_n, S_n^2)$ 是 $\theta = (\mu, \sigma^2)$ 的充分统计量。
  - ✓ 然而，若随机样本并非来自正态分布，那么 $(\bar{X}_n, S_n^2)$ 就可能不是充分统计量。
- 换言之，一个统计量 $T(\boldsymbol{X}^n)$ 是否为充分统计量通常依赖于具体的总体分布，在某些总体分布下是 θ 的充分统计量，在其他总体分布下则可能不是充分统计量。

◆ **问题 6.6**

能否举出一个总体分布的例子，使得$(\bar{X}_n, S_n^2)$ 不是 $\theta = (\mu, \sigma^2)$ 的充分统计量?

定理 6.11

[不变性原理 (Invariance Principle)]：若 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量，则任意一一对应的函数 $R(\boldsymbol{X}^n) = r[T(\boldsymbol{X}^n)]$ 也是 θ 的充分统计量，同时也是变换参数 $r(\theta)$ 的充分统计量。

证明：

- 因为 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量，存在函数 $g(\cdot,\cdot)$ 和 $h(\cdot)$，使得随机样本 $\boldsymbol{X}^n$ 的联合 PMF/PDF 可写为

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) = g[T(\boldsymbol{x}^n), \theta]h(\boldsymbol{x}^n)$$

证明（Cont.）：

- 又因为 $r(\cdot)$ 是一一映射，其反函数 $r^{-1}(\cdot)$ 存在，并满足 $T(\boldsymbol{x}^n) = r^{-1}[R(\boldsymbol{x}^n)]$。则有

$$\begin{aligned} f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) &= g\{r^{-1}[R(\boldsymbol{x}^n)], \theta\} h(\boldsymbol{x}^n) \\ &= \tilde{g}[R(\boldsymbol{x}^n), \theta] h(\boldsymbol{x}^n) \end{aligned}$$

其中，变换函数 $\tilde{g}(\cdot, \theta) = g[r^{-1}(\cdot), \theta]$ 依赖于参数 θ。

- 根据充分统计量的定义，$R(\boldsymbol{X}^n)$ 是 θ 的充分统计量。

证明（Cont.）：

- 类似地，因为 $\theta = r^{-1}[r(\theta)] = r^{-1}(\beta)$，其中 $\beta = r(\theta)$ 为变换参数，则有

$$\begin{aligned} f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) &= g\{r^{-1}[R(\boldsymbol{x}^n)], r^{-1}(\beta)\}h(\boldsymbol{x}^n) \\ &= g^*[R(\boldsymbol{x}^n), \beta]h(\boldsymbol{x}^n) \end{aligned}$$

  其中 $g^*(\cdot, \beta) = g[r^{-1}(\cdot), r^{-1}(\beta)]$ 是参数 β 的函数。

- 因此，$R(\boldsymbol{X}^n)$ 也是 β 的充分统计量。

证毕。

定义 6.9

[指数分布族 (Exponential Family)]：概率分布族称为指数分布族，若其总体 PMF/PDF 可表示为

$$f(x,\theta)=h(x)c(\theta)e^{\sum_{j=1}^{k}w_j(\theta)t_j(x)}$$

- 第四章介绍的绝大多数重要分布——包括离散分布和连续分布——都属于指数分布族。
- 正态分布 $N(\mu,\sigma^2)$ 即为一例，其 PDF 为

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

- 其中

$$h(x) = 1$$

$$c(\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\mu^2}{2\sigma^2}}$$

$$w_1(\theta) = -\frac{1}{2\sigma^2}$$

$$w_2(\theta) = \frac{\mu}{\sigma^2}$$

$$t_1(x) = x^2$$

$$t_2(x) = x$$

定理 6.12

令 $\boldsymbol{X}^n = (X_1, \cdots, X_n)$ 为来自总体 PMF/PDF 为 $f(x, \theta)$ 的 IID 随机样本。若

$$f(x, \theta) = h(x)c(\theta)e^{\sum_{j=1}^{k} w_j(\theta)t_j(x)}$$

则 $k \times 1$ 统计向量

$$T(\boldsymbol{X}^n) = [\sum_{i=1}^{n} t_1(X_i), \cdots, \sum_{i=1}^{n} t_k(X_i)]'$$

是 θ 的充分统计量。

证明：

- 留作练习题。

定义 6.10

[最小充分统计量 (Minimal Sufficient Statistic)]：若对任何其他充分统计量 $R(\boldsymbol{X}^n)$，$T(\boldsymbol{X}^n)$ 总是 $R(\boldsymbol{X}^n)$ 的函数，即对任意充分统计量 $R(\boldsymbol{X}^n)$，总存在一个函数 $r(\cdot)$ 满足 $T(\boldsymbol{X}^n) = r[R(\boldsymbol{X}^n)]$，则称充分统计量 $T(\boldsymbol{X}^n)$ 为参数 θ 的最小充分统计量。

- 参数 θ 的所有充分统计量都包含了与 θ 相关的所有样本信息，但最小充分统计量在 θ 的所有充分统计量中实现了对数据集的最大可能概括。
- 为什么呢?

- 为说明这一点，假设 $T(\boldsymbol{X}^n)=r[R(\boldsymbol{X}^n)]$，且 $t=r(\tau)$。定义 $\boldsymbol{X}^n$ 的样本空间中样本点的两个子集：

$$A_n(\tau)=\{\boldsymbol{x}^n: R(\boldsymbol{x}^n)=\tau\}$$
$$B_n(t)=\{\boldsymbol{x}^n: T(\boldsymbol{x}^n)=t\}$$
$$=\{\boldsymbol{x}^n: r[R(\boldsymbol{x}^n)]=r(\tau)\}$$

  - ✓ $A_n(\tau)$ 以 τ 为参数，
  - ✓ $B_n(t)$ 以 t 为参数，其中 $t=r(\tau)$。

  则 $A_n(\tau)\subseteq B_n(t)$，因为 $R(\boldsymbol{x}^n)=\tau$ 可推出 $T(\boldsymbol{x}^n)=r[R(\boldsymbol{x}^n)]=r(\tau)=t$，但 $T(\boldsymbol{x}^n)=t$ 无法推出 $R(\boldsymbol{x}^n)=\tau$。

- 因此，与 $R(\boldsymbol{x}^n)=\tau$ 相比，$T(\boldsymbol{x}^n)=t$ 概括的样本信息集 $B_n(t)$ 是一个更大的集合。

◆ **问题 6.7**

如何求得 θ 的最小充分统计量?

- 以下定理提供了一种检验统计量 $T(\boldsymbol{X}^n)$ 是否为参数 θ 的最小充分统计量的便捷方法。

定理 6.13

- 令 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)$ 为随机样本 $\boldsymbol{X}^n$ 的 PMF/PDF。
- 假设对随机样本 $\boldsymbol{X}^n$ 的样本空间中的任意两个样本点 $\boldsymbol{x}^n$ 与 $\boldsymbol{y}^n$，存在函数 $T(\boldsymbol{x}^n)$，当且仅当 $T(\boldsymbol{x}^n) = T(\boldsymbol{y}^n)$ 时，联合 PMF/PDF 之比 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)/f_{\boldsymbol{X}^n}(\boldsymbol{y}^n, \theta)$ 为参数 θ 的常函数 (即不依赖于 θ)。
- 则 $T(\boldsymbol{X}^n)$ 为 θ 的最小充分统计量。

证明：

(1) 首先证明在给定条件下，$T(\boldsymbol{X}^n)$ 是 θ 的充分统计量。

- 定义 $A(t)=\{\boldsymbol{x}^n: T(\boldsymbol{x}^n)=t\}$ 为随机样本 $\boldsymbol{X}^n$ 的样本空间中的样本点的一个集合。
- 对每一个 $A(t)$，选择并固定一个元素 $\boldsymbol{x}_t^n \in A(t)$。换言之，对任意样本点 $\boldsymbol{x}^n \in A(t)$，令 $\boldsymbol{x}_t^n$ 为与 $\boldsymbol{x}^n$ 在同一集合 $A(t)$ 中的一个固定元素。
- 因为 $\boldsymbol{x}^n$ 和 $\boldsymbol{x}_t^n$ 在同一个集合 $A(t)$ 中，故 $T(\boldsymbol{x}^n)=T(\boldsymbol{x}_t^n)$。
- 因此，在满足定理所给定的假设条件下，有 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)/f_{\boldsymbol{X}^n}(\boldsymbol{x}_t^n,\theta)$ 为 θ 的常函数（即不依赖于 θ）。

证明 (Cont.)：

- 可定义函数 $h(\boldsymbol{x}^n) = f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)/f_{\boldsymbol{X}^n}(\boldsymbol{x}_t^n, \theta)$，这里 $h(\boldsymbol{x}^n)$ 不依赖于 θ，且只是 $\boldsymbol{x}^n$ 的函数（注意 $t = T(\boldsymbol{x}^n)$ 是 $\boldsymbol{x}^n$ 的函数，因此 $\boldsymbol{x}_t^n$ 也是 $\boldsymbol{x}^n$ 的函数）。
- 同时，定义函数 $g(t, \theta) = f_{\boldsymbol{X}^n}(\boldsymbol{x}_t^n, \theta)$。则有

$$
\begin{aligned}
f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) &= \frac{f_{\boldsymbol{X}^n}(\boldsymbol{x}_t^n, \theta) f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta)}{f_{\boldsymbol{X}^n}\left(\boldsymbol{x}_t^n, \theta\right)} \\
&= f_{\boldsymbol{X}^n}(\boldsymbol{x}_t^n, \theta) h(\boldsymbol{x}^n) \\
&= g(t, \theta) h(\boldsymbol{x}^n) \\
&= g[T(\boldsymbol{x}^n), \theta] h(\boldsymbol{x}^n) \quad [\text{使用} t = T(\boldsymbol{x}^n)]
\end{aligned}
$$

- 根据因子分解定理（定理 6.10），$T(\boldsymbol{X}^n)$ 是 θ 的充分统计量。

证明 (Cont.):

(2) 现在证明在给定假设条件下，$T(\boldsymbol{X}^n)$ 是最小充分统计量。令 $\tilde{T}(\boldsymbol{X}^n)$ 为 θ 的另一个充分统计量。

- 由因子分解定理 (定理 6.10) 可知，存在函数 $\tilde{g}(\cdot,\cdot)$ 和 $\tilde{h}(\cdot)$ 满足 $f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)=\tilde{g}[\tilde{T}(\boldsymbol{x}^n),\theta]\tilde{h}(\boldsymbol{x}^n)$。令 $\boldsymbol{x}^n$ 和 $\mathbf{y}^n$ 为 $\boldsymbol{X}^n$ 的样本空间的任意两个样本点且 $\tilde{T}(\boldsymbol{x}^n)=\tilde{T}(\mathbf{y}^n)$，则

$$\frac{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)}{f_{\boldsymbol{X}^n}(\mathbf{y}^n,\theta)}=\frac{\tilde{g}\left[\tilde{T}(\boldsymbol{x}^n),\theta\right]\tilde{h}(\boldsymbol{x}^n)}{\tilde{g}\left[\tilde{T}(\mathbf{y}^n),\theta\right]\tilde{h}(\mathbf{y}^n)}=\frac{\tilde{h}(\boldsymbol{x}^n)}{\tilde{h}(\mathbf{y}^n)}$$

不依赖于 θ。

证明 (Cont.)：

- 因为该比例不依赖于 θ，定理 6.13 所给定的充要假设条件表明 $T(\boldsymbol{x}^n) = T(\mathbf{y}^n)$。换言之，从 $\tilde{T}(\boldsymbol{x}^n) = \tilde{T}(\mathbf{y}^n)$ 可推出 $T(\boldsymbol{x}^n) = T(\mathbf{y}^n)$，故

$$\{\mathbf{y}^n : \tilde{T}(\mathbf{y}^n) = \tilde{T}(\boldsymbol{x}^n)\} \subseteq \{\mathbf{y}^n : T(\mathbf{y}^n) = T(\boldsymbol{x}^n)\}$$

- 因此，$T(\boldsymbol{x}^n)$ 为最小充分统计量。

证毕。

例 6.13：

- 令 $\boldsymbol{X}^n$ 为来自总体 $N(\mu,\sigma^2)$ 分布的 IID 随机样本，其中参数 μ 和 σ^2 未知。
- 令 $\boldsymbol{x}^n$ 和 $\boldsymbol{y}^n$ 表示 $\boldsymbol{X}^n$ 的样本空间中的任意两个样本点，并令 $(\bar{x}_n, s_X^2)$ 和 $(\bar{y}_n, s_Y^2)$ 分别为 $\boldsymbol{x}^n$ 和 $\boldsymbol{y}^n$ 的样本均值和样本方差。
- 当且仅当 $(\bar{x}_n, s_X^2) = (\bar{y}_n, s_y^2)$，有

$$\frac{f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta)}{f_{\boldsymbol{X}^n}(\boldsymbol{y}^n,\theta)} = \frac{(2\pi\sigma^2)^{-n/2} e^{-[n(\bar{x}_n-\mu)^2+(n-1)s_X^2]/2\sigma^2}}{(2\pi\sigma^2)^{-n/2} e^{-[n(\bar{y}_n-\mu)^2+(n-1)s_Y^2]/2\sigma^2}} = 1$$

不依赖于 θ。

- 因此，$(\bar{x}_n, s_X^2)$ 为 (μ,σ^2) 的最小充分统计量。

目　录

- **统计分析的基本思想：**利用子集或样本信息推断数据生成过程的信息。
- 本章介绍了统计抽样理论的基本概念与思想，**相关概念**包括：

总体	随机样本	数据集	统计量	参数	统计推断

- 详细分析了**两个重要的统计量**：
  - ✓ **样本均值估计量**
  - ✓ **样本方差估计量**

- 在独立同分布正态随机样本假设下构建了**经典有限样本抽样分布理论**。
  - ✓ 该有限样本理论突显了 t- **分布**和 $\mathcal{F}-$ **分布**在统计推断中的重要性。
- 最后，介绍了**充分统计量**的概念和思想，并对其在数据简化中的作用进行了讨论。
  - ✓ 充分性原则很好概括了统计分析的本质思想，即如何最有效地概括观测数据，以推断总体分布或总体分布的参数。

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 !