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

第九章 假设检验

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

摘要

- 假设检验是统计推断的两个重要目的之一。
- 本章将介绍假设检验的
  - ✓ 基本概念
  - ✓ 基本思想
  - ✓ 基本理论
- 并讨论参数假设检验中的三大基本检验方法：
  - ✓ 沃尔德检验 (Wald test)
  - ✓ 拉格朗日乘子检验 (Lagrange multiplier test)
  - ✓ 似然比检验 (likelihood ratio test)
- 同时，指出了经济假说和参数假设之间的联系与区别。

目　录

例 9.1 [教育回报 (Returns to Education)]:

- 令参数 θ 测量当其他因素不变时，增加一年教育时间所带来的时薪率的变化。
- 劳动经济学家感兴趣的是，在控制其他因素的情况下，检验教育收益是否为零，即 θ 是否等于零。

$$\mathbb{H}_0\colon \theta = 0$$

定义 9.1 [假设 (Hypothesis)]

- 假设是关于总体分布或总体分布某些特征的表述。
- 假设检验问题中有两个互补的假设，称为**原假设** (null hypothesis) 和**备择假设** (alternative hypothesis)，分别用 $\mathbb{H}_0$ 和 $\mathbb{H}_A$ 表示。
  - ✓ 原假设 $\mathbb{H}_0$ 是关于总体分布或总体分布中某些特征的陈述，
  - ✓ 备择假设 $\mathbb{H}_A$ 则是对原假设的否定。

- **检验原假设的目的**就是：基于由总体分布生成的观测数据集 $\boldsymbol{x}^n$，判断两个互补假设哪个为真。
- 假定一个随机样本 $\boldsymbol{X}^n$ 由总体分布 $f(x,\theta)$ 生成，参数 $\theta \in \Theta$ 的真实值未知，Θ 是已知有限维数的参数空间。
- 假设检验中，参数空间 Θ 被划分为两个**互斥且完备**的子集 Θ_0 和 Θ_A，即 $\Theta_0 \cap \Theta_A = \emptyset$ 且 $\Theta_0 \cup \Theta_A = \Theta$。

- 所关注的问题是确定参数 θ 的真实值属于这两个子集的哪一个，即基于一个观测数据集 $\boldsymbol{x}^n$，在如下两个假设中二选一
  - ✓ $\mathbb{H}_0: \theta \in \Theta_0$
  - ✓ $\mathbb{H}_A: \theta \in \Theta_A$
- $\mathbb{H}_0$ 称为“原 (null)”假设，因为其代表的内容通常是“没有效果”或“没有关系”。

例 9.2：

- 令 θ 表示某工业品中次品所占比例。
- 考虑检验 $\mathbb{H}_0: \theta \le \theta_0$ 和 $\mathbb{H}_A: \theta > \theta_0$，其中 θ_0 为已知常数，是次品的可接受最大比例。
- 原假设指出次品所占比例不超过某一事先设定的不合格临界值。该假设是统计质量控制的基本思想。

例 9.3 [检验规模报酬不变假说 (Testing Constant Returns to Scale Hypothesis)]:

- 生产函数

$$Y = F(L, K)$$

给出了当投入资本 K 和劳动 L 时可获得多少产出 Y。

- 若产出与投入的增幅相同，即对所有常数 $\lambda > 0$,

$$\lambda F(L, K) = F(\lambda L, \lambda K)$$

则生产技术称为规模报酬不变。

- 假设生产函数形式为柯布-道格拉斯函数

$$Y = AK^{\alpha}L^{\beta}$$

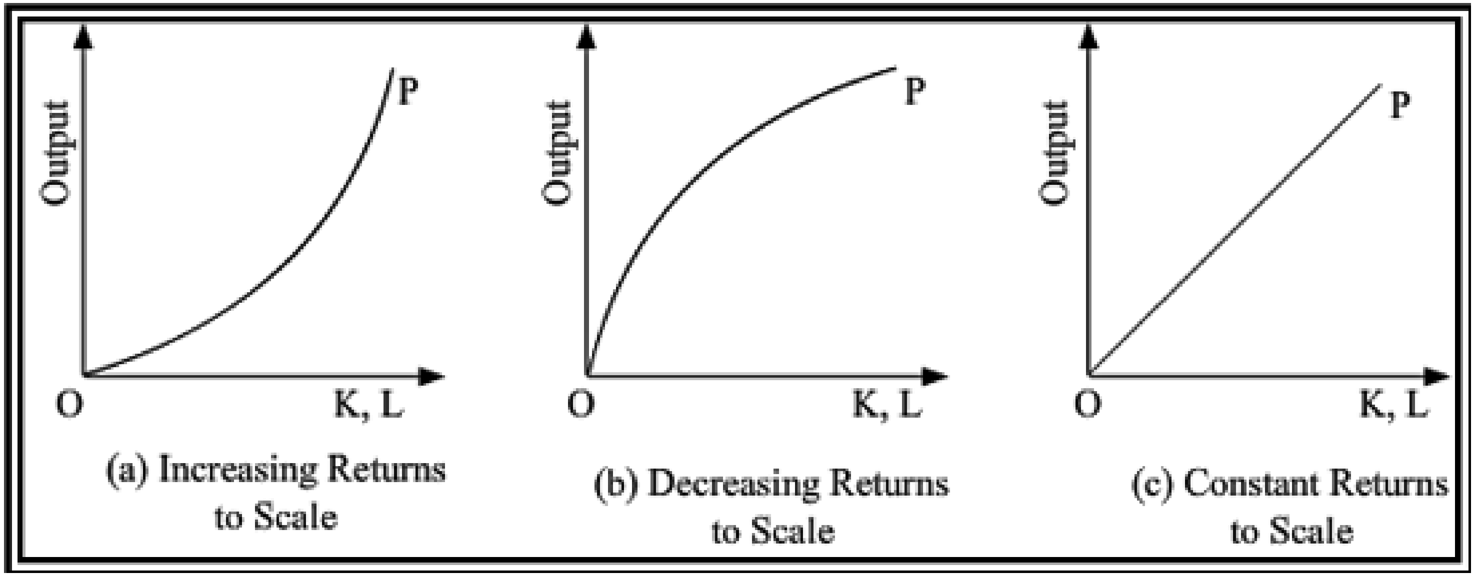
其中 A 为常数，$\theta = (\alpha, \beta)$ 为参数向量,

例 9.3 (Cont.):

- 则**规模报酬不变**假说可表达为

$$\mathbb{H}_0: \alpha + \beta = 1$$

- 备择假设 $\mathbb{H}_A$: $\alpha + \beta \neq 1$ 包括两种情况:
  - ✓ $\alpha + \beta > 1$：对应**规模报酬递增**，
  - ✓ $\alpha + \beta < 1$：对应**规模报酬递减**。



定义 9.2 [简单假设与复合假设 (Simple Hypothesis Versus Composite Hypothesis)]

- 当且仅当假设中只有一个总体分布时，该假设称为**简单假设**。
- 若假设包含了多个总体分布，则称其为**复合假设**。

- 对参数分布模型 $f(x,\theta)$，参数 θ 的每一个值对应一个总体分布，θ 的不同参数值对应不同的总体分布。
- 在例 9.1 中，原假设 $\mathbb{H}_0$ 仅包含一个参数值，因此 $\mathbb{H}_0$ 是简单假设。
- 相反，例 9.2 和例 9.3 的原假设包括不止一个参数值，因此是复合假设。

定义 9.3 [假设检验 (Hypothesis Testing)]

假设检验是一种统计决策规则，它设定了

- 对什么样本值 x^n，决定无法拒绝 $\mathbb{H}_0$ 为真；
- 对什么样本值 x^n，决定拒绝 $\mathbb{H}_0$ 而接受 $\mathbb{H}_A$ 为真。

定义 9.4 [临界域或拒绝域 (Critical Region; Rejection Region)]

- 随机样本 X^n 的样本空间中那些将拒绝 $\mathbb{H}_0$ 的样本点的集合 $\mathbb{C}$ 称为**拒绝域** (Rejection Region) 或**临界域** (Critical Region)。
- 拒绝域的补集称为**接受域** (Acceptance Region)。

- **假设检验的标准方法**是选择一个统计量 $T(\boldsymbol{X}^n)$，并用它将 $\boldsymbol{X}^n$ 的样本空间划分为互斥且完全穷尽的两个区域

$$\mathbb{A}_n(c) = \{\boldsymbol{x}^n : T(\boldsymbol{x}^n) \le c\}$$

和

$$\mathbb{C}_n(c) = \{\boldsymbol{x}^n : T(\boldsymbol{x}^n) > c\}$$

其中 c 是某预先设定的常数。

✓ 第一个区域 $\mathbb{A}_n(c)$ 是**接受域**，

✓ 第二个区域 $\mathbb{C}_n(c)$ 是**拒绝域**。

✓ 分界点 c 称为**临界值**，

✓ 而 $T(\boldsymbol{X}^n)$ 称为**检验统计量**。

例 9.4：

- 假设 $\boldsymbol{X}^n$ 为 IID $N(\mu,\sigma^2)$ 随机样本，其中 μ 未知，但 σ^2 已知。
- 考虑检验 $\mathbb{H}_0:\mu=\mu_0$ 和 $\mathbb{H}_A:\mu\neq\mu_0$，其中 μ_0 为已知常数。这里，$\Theta_0=\{\mu_0\}$ 只包含一个参数值 μ_0，而 Θ_A 包含实数线上所有除 μ_0 以外的参数值。
- 为检验原假设 $\mathbb{H}_0:\mu=\mu_0$，考虑如下的检验统计量

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

- 由假设可知，σ 为已知常数。在原假设 $\mathbb{H}_0:\mu=\mu_0$ 下，有

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

例 9.4（Cont.）：

- 在备择假设 $\mathbb{H}_A: \mu \neq \mu_0$ 下，

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

当 $n \to \infty$ 时， $T(\boldsymbol{X}^n)$ 以概率趋于 1 发散到无穷。

- 因此，
  - ✓ 若绝对值 $|T(\boldsymbol{X}^n)|$ 很大则接受 $\mathbb{H}_A$，
  - ✓ 若 $|T(\boldsymbol{X}^n)|$ 不大则无法拒绝 $\mathbb{H}_0$。

例 9.4 (Cont.):

- 那么， $|T(\boldsymbol{X}^n)|$ 的值需要多大才能判定为“大”？
  - ✓ 这取决于 $T(\boldsymbol{X}^n)$ 在原假设 $\mathbb{H}_0$ 下的抽样分布 $N(0,1)$。
- 具体而言，令 $c = z_{\alpha/2}$，其中 $z_{\alpha/2}$ 是 $N(0,1)$ 在 $\alpha/2 \in (0,1/2)$ 显著水平上的右侧临界值，即 $P(Z > z_{\alpha/2}) = \alpha/2$，这里 $Z \sim N(0,1)$。
- 定义**接受域**和**拒绝域**如下

$$\mathbb{A}_n(c) = \left\{\boldsymbol{X}^n : \left|\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right| \leq z_{\alpha/2}\right\}$$

$$\mathbb{C}_n(c) = \left\{\boldsymbol{X}^n : \left|\frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}\right\}$$

例 9.4 (Cont.):

- 从而得到如下决策规则:
  - ✓ 在显著水平 α 上无法拒绝 $\mathbb{H}_0: \mu = \mu_0$，若
  $$\boldsymbol{x}^n \in \mathbb{A}_n(c)$$
  - ✓ 在显著水平 α 上拒绝 $\mathbb{H}_0: \mu = \mu_0$，若
  $$\boldsymbol{x}^n \in \mathbb{C}_n(c)$$

正态检验统计量 $T(X^n)$ 的拒绝域和接受域

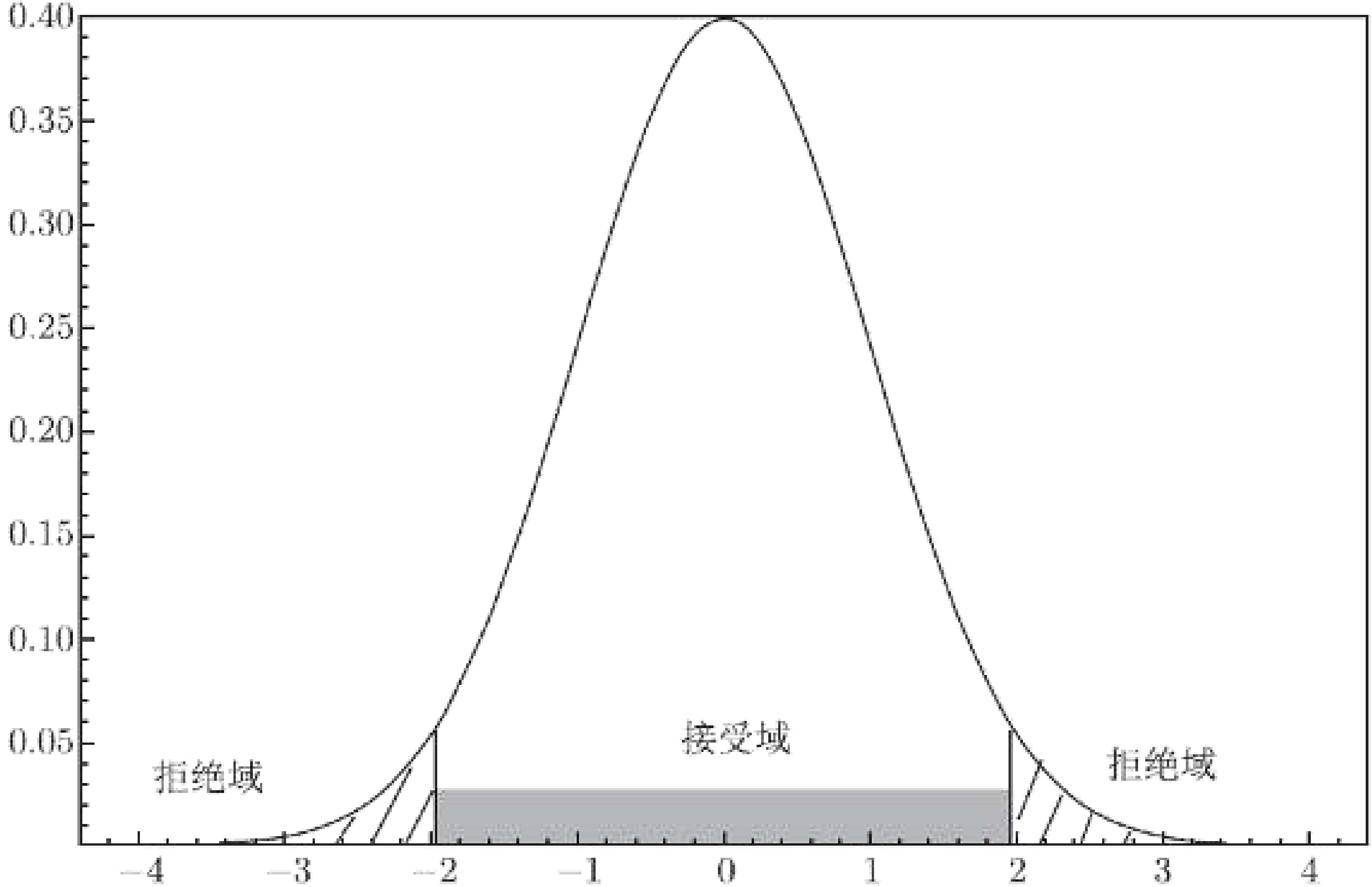


图 9.1 ：拒绝域与接受域

定义 9.5 [检验功效 (Power of Test)]

若 $\mathbb{C}$ 是检验原假设 $\mathbb{H}_0 : \theta \in \Theta_0$ 的拒绝域，则函数 $\pi(\theta) = P_\theta(\boldsymbol{X}^n \in \mathbb{C})$ 称为拒绝原假设 $\mathbb{H}_0$ 的检验功效，其中 $P_\theta(\cdot)$ 是当随机样本 $\boldsymbol{X}^n$ 服从分布 $f_{X^n}(\boldsymbol{x}^n, \theta)$ 时的概率测度。

- 检验功效函数 $\pi(\theta)$ 是拒绝 $\mathbb{H}_0$ 的概率。在例 9.4 中，正态检验统计量 $T(\boldsymbol{X}^n) = \sqrt{n}(\bar{X}_n - \mu_0)/\sigma$ 的功效为

$$\pi(\mu) = P\left(\left|\frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}\right| > z_{\alpha/2}\right)$$

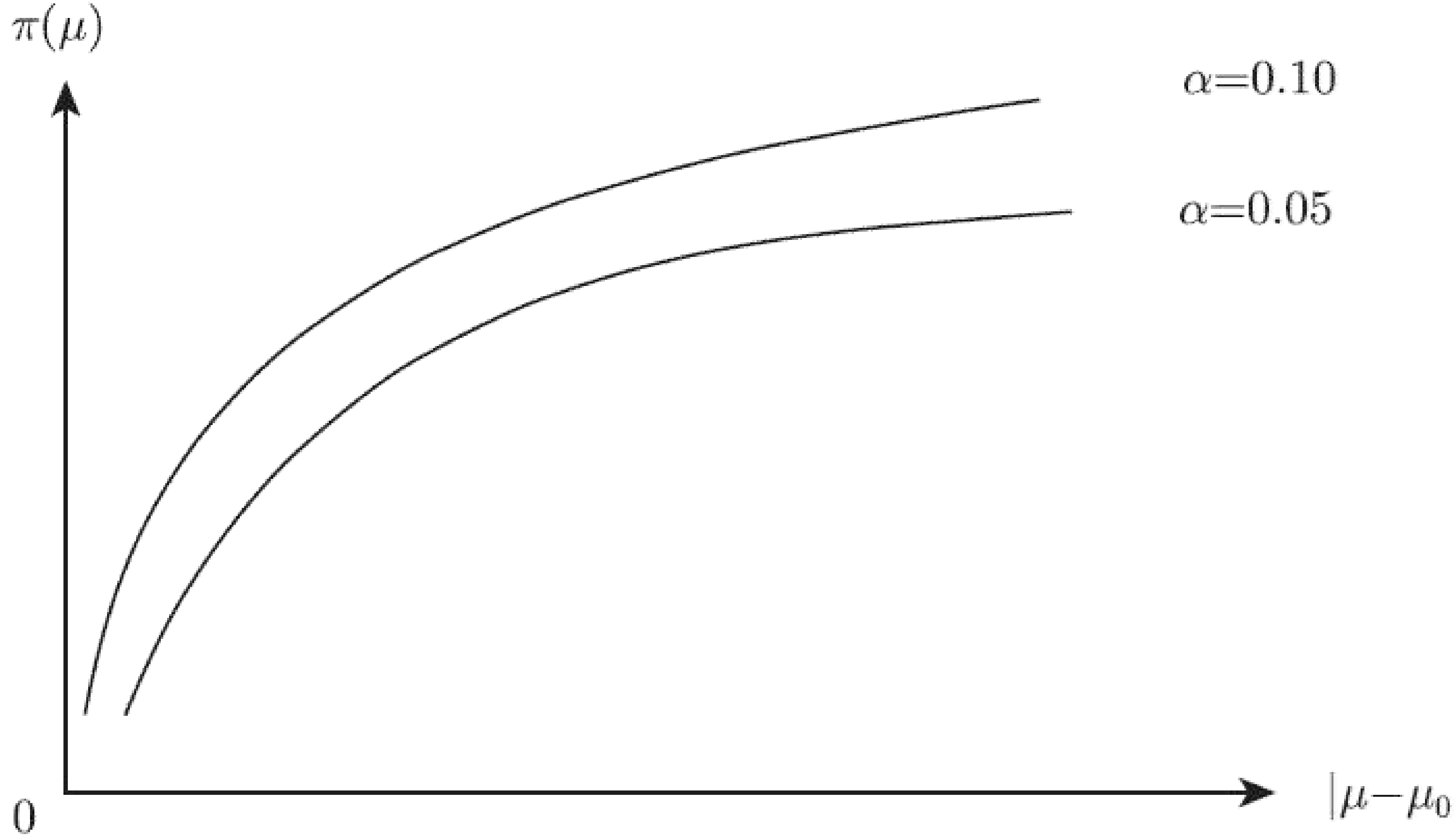


图 9.2：检验统计量 $\dfrac{\bar{X}_n-\mu_0}{\sigma/\sqrt{n}}$ 在 10% 和 5% 显著水平上的功效函数

定义 9.6 [第一类错误和第二类错误 (Type I and Type II Errors)]

- 若 $\mathbb{H}_0: \theta \in \Theta_0$ 成立，但观测数据 $\boldsymbol{x}^n$ 落在临界域 $\mathbb{C}$ 中，则犯了第一类错误。犯**第一类错误**的概率为

$$\alpha(\theta) \equiv P_\theta(\boldsymbol{X}^n \in \mathbb{C} \mid \mathbb{H}_0)$$

- 若 $\mathbb{H}_A: \theta \in \Theta_0^c$ 成立但观测数据 $\boldsymbol{x}^n$ 落在接受域 $\mathbb{A}$ 内，则犯了第二类错误。犯**第二类错误**的概率为

$$\begin{aligned}\beta(\theta) &\equiv P_\theta(\boldsymbol{X}^n \in \mathbb{A} \mid \mathbb{H}_A) \\ &= 1 - P_\theta(\boldsymbol{X}^n \in \mathbb{C} \mid \mathbb{H}_A)\end{aligned}$$

第一类错误和第二类错误

- 在 $\mathbb{H}_0: \theta \in \Theta_0$ 下，**功效函数** (power function) $\pi(\theta)$ 是犯第一类错误的概率，即错误拒绝正确原假设的概率。
  - ✓ 第一类错误无法避免，因为在原假设 $\mathbb{H}_0$ 下，检验统计量 $T(\boldsymbol{X}^n)$ 仍有一定概率 (尽管较小) 取较大值。
- 在 $\mathbb{H}_A: \theta \in \Theta_0^c$ 下，功效函数 $\pi(\theta)$ 是拒绝错误原假设的概率，其值等于 $1-\beta(\theta)$，其中 $\beta(\theta)$ 是犯第二类错误的概率，即无法拒绝错误原假设的概率。
  - ✓ 有多种原因可能导致第二类错误。例如，样本容量 n 较小使得检验统计量 $T(\boldsymbol{X}^n)$ 以非零概率取较小值。

无偏检验 (Unbiased test)

- 若

$$P[T(\boldsymbol{X}^n) > c \,|\mathbb{H}_A] > P[T(\boldsymbol{X}^n) > c|\mathbb{H}_0\,],$$

则**检验为无偏**。

- 即当 $\mathbb{H}_0$ 为假时，拒绝原假设 $\mathbb{H}_0$ 的概率严格大于当其为真时被拒绝的概率。

两类错误之间的权衡取舍

- 理想情况下，当然希望对所有 $\theta \in \Theta_0$ 统计检验的功效函数 $\pi(\theta)$ 均等于 0，而对所有 $\theta \in \Theta_A$，则功效函数均等于 1。
- 然而，对任何统计检验，给定任意样本容量 n，都存在第一类错误和第二类错误之间的权衡取舍。
- 对任意给定的 n，若临界域 C 变小，则犯第一类错误的概率将减小，但犯第二类错误的概率上升。反之，若临界域 $\mathbb{C}$ 增大，则第二类错误 $\beta(\theta)$ 减少，但第一类错误 $\alpha(\theta)$ 增加。
- 给定一个检验，唯一可同时减少两类错误的办法是增加样本容量 n。

- 通常，假设检验是通过犯错误的概率进行评价与比较。因为上述两类错误的概率在许多情况下存在负相关，故假设检验的通常方法是对 Θ_0 中的所有 θ，以某个值 $\alpha \in (0,1)$ 限定第一类错误的概率，同时寻找一个检验使得对 Θ_A 中的所有 θ，该检验犯第二类错误的概率最小。

一致最大功效检验 (Uniformly Most Powerful Test)

- 具体而言，需要找一个对所有 $\theta \in \Theta_0$，满足

$$P[T(\boldsymbol{X}^n) > c|\mathbb{H}_0] \leq \alpha$$

的检验统计量 $T(\boldsymbol{X}^n)$，且对所有 $\theta \in \Theta_A$，满足

$$P[T(\boldsymbol{X}^n) > c\ |\mathbb{H}_A] \geq P[G(\boldsymbol{X}^n) > c\ |\mathbb{H}_A],$$

这里 $G(\boldsymbol{X}^n)$ 为其他任意检验统计量。

- 换言之，需要找一个有最大功效的检验统计量 $T(\boldsymbol{X}^n)$，同时其第一类错误可控。
- 具有这些性质的检验 $T(\boldsymbol{X}^n)$ 称为**一致最大功效检验**，所谓“一致”是指该检验对所有 $\theta \in \Theta_A$ 均有最大功效。

定义 9.7 [一致最大功效检验 (Uniformly Most Powerful Test)]

- 令 $\mathbb{T}$ 为一族关于 $\mathbb{H}_0: \theta \in \Theta_0$ 和 $\mathbb{H}_A: \theta \in \Theta_A$ 的检验集合，且 $\pi(\theta)$ 为某一检验 $T(\boldsymbol{X}^n) \in \mathbb{T}$ 的功效函数。
- 若对所有 $\theta \in \Theta_A$，$\pi(\theta) \geq \tilde{\pi}(\theta)$，其中 $\tilde{\pi}(\theta)$ 是 $\mathbb{T}$ 集合中任意其他检验 $G(\boldsymbol{X}^n)$ 的功效函数。
- 则检验 $T(\boldsymbol{X}^n)$ 称为 $\mathbb{T}$ 中的一致最大功效检验。

检验水平 (level) 与尺度 (size)

- 假设对检验统计量 $T(\boldsymbol{X}^n)$，有

$$P[T(\boldsymbol{X}^n) > c|\mathbb{H}_0] \le \alpha$$

则 α 的值是检验统计量 $T(\boldsymbol{X}^n)$ 犯第一类错误的最大值，称为**检验水平 (the level of the test)**。

- 若 $T(\boldsymbol{X}^n)$ 的水平为 α 且

$$P[T(\boldsymbol{X}^n) > c|\mathbb{H}_0] = \alpha$$

则该检验称为**尺度 (size)** 为 α 的检验。

- 显然，一族水平为 α 的统计检验包含了尺度为 α 的检验。

检验水平 (level) 与尺度 (size)（Cont.)

- 尺度为 α 的检验精确地给出了犯第一类错误的概率。通常，用 $\mathbb{T}$ 表示一族具有相同显著性水平或相同尺度 α 的检验。
- 但在一些复杂的检验情形中，可能计算上会比较困难，甚至无法构造尺度为 α 的假设。此时一般使用水平为 α 的检验。
- 实际应用中，通常事先设定某一检验水平，如设定 $\alpha = 0.01, 0.05, 0.10$。确定了检验水平，也就控制了第一类错误发生的概率。

例 9.5 [检验有效市场假说 (Efficient Market Hypothesis, EMH)]:

- 假设
  - ✓ R_t 是某投资组合在第 t 期的收益率,
  - ✓ $I_{t-1} = (R_{t-1}, R_{t-2}, \cdots)$ 为第 $t-1$ 期包含的所有历史收益率的信息集。
- 若

$$E(R_t \mid I_{t-1}) = E(R_t)$$

则称资产市场为信息弱式有效 (informationally weakly efficient)，即资产收益率的历史信息不能预测未来的资产收益率。

例 9.5 (Cont.):

- 为检验这一经济假说，可考虑如下线性自回归模型

$$R_t = \alpha + \sum_{j=1}^{k} \beta_j R_{t-j} + \varepsilon_t$$

其中 ε_t 是随机扰动项。

- 在有效市场假说下，有

$$\mathbb{H}_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$$

任何斜率系数 β_j，$j \in \{1, \cdots, k\}$ 不等于零都是拒绝 EMH 的证据。

例 9.5 (Cont.)：

- 但是，若无法拒绝统计假设 $\mathbb{H}_0$，这并不意味着原始的经济假说——有效市场假说成立，因为线性自回归模型仅是多种检验资产收益率的历史信息对未来资产收益率是否有预测能力的方法之一。
- 例如，历史信息的预测性可能以非线性的形式出现。
- 因此，有效市场假说和统计假设 $\mathbb{H}_0$ 之间存在一定差距。正因为这一差距的存在，若无法拒绝 $\mathbb{H}_0$，只能称未发现推翻有效市场假说的证据，而无法得出有效市场假说成立的结论。

目　录

定理 9.1 [内曼-皮尔逊引理 (Neyman-Pearson Lemma)]

- 考虑检验一个简单的原假设 $\mathbb{H}_0: \theta = \theta_0$ 和一个简单的备择假设 $\mathbb{H}_A: \theta = \theta_1$，其中对应于 $\theta_i\ (i = 0, 1)$ 的随机样本 $\boldsymbol{X}^n$ 的 PMF/PDF 为 $f_{X^n}(\boldsymbol{x}^n, \theta_i)$。
- 给定某个常数 $c \geq 0$，定义一个检验的**拒绝域** $\mathbb{C}_n(c)$ 和**接受域** $\mathbb{A}_n(c)$ 分别为

(a)

$$\mathbb{C}_n(c) = \left\{\boldsymbol{x}^n: \frac{f_{X^n}(\boldsymbol{x}^n, \theta_1)}{f_{X^n}(\boldsymbol{x}^n, \theta_0)} > c\right\}$$

和

$$\mathbb{A}_n(c) = \left\{\boldsymbol{x}^n: \frac{f_{X^n}(\boldsymbol{x}^n, \theta_1)}{f_{X^n}(\boldsymbol{x}^n, \theta_0)} \leq c\right\}$$

定理 9.1 (Cont.)

(b)

$$P[\boldsymbol{X}^n \in \mathbb{C}_n(c) \mid \mathbb{H}_0] = \alpha$$

则

(1) [充分性] 满足上述条件 (a) 和 (b) 的任意检验是水平为 α 的一致最大功效检验;

(2) [必要性] 若存在一个检验，当 $c > 0$ 时满足上述条件，则每个水平为 α 的一致最大功效检验均为尺度 α 的检验，即满足条件 (b)；且每个水平为 α 的一致最大功效检验都满足上述条件 (a)，除了可能一个有 $P(\boldsymbol{X}^n \in \mathbb{A}|\mathbb{H}_0) = P(\boldsymbol{X}^n \in \mathbb{A}|\mathbb{H}_A) = 0$ 的零概率集合 $\mathbb{A}$ 之外。

证明：

- 注意到第一类错误的概率

$$P[\boldsymbol{X}^n \in \mathbb{C}_n(c) \mid \mathbb{H}_0] = E\{\mathbf{1}[\boldsymbol{X}^n \in \mathbb{C}_n(c)] \mid \mathbb{H}_0\}$$
$$= \int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] f(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n$$

其中 $\mathbf{1}(\cdot)$ 是指示函数，若其内部的条件成立则取值 1，否则取值 0。

- **(1)** 首先证明，一个满足条件 (a) 和 (b) 的检验 (比如 $T(\boldsymbol{X}^n)$) 具有一致最大功效。
- 假设有另一检验 (比如 $T_1(\boldsymbol{X}^n)$) 满足 $E\{\mathbf{1}[\boldsymbol{X}^n \in \mathbb{C}_{1n}] | \mathbb{H}_0\} \le \alpha$ (注意，$T_1(\boldsymbol{X}^n)$ 不一定是似然比检验)。

证明 (Cont.)：

- 以下将证明 $T(\boldsymbol{X}^n)$ 较 $T_1(\boldsymbol{X}^n)$ 的功效更大。
- 若 $\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] > \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]$，则样本点 $\boldsymbol{x}^n$ 处于检验 $T(\boldsymbol{X}^n)$ 的拒绝域 $\mathbb{C}_n(c)$ 内，因而有
$$f_{X^n}(\boldsymbol{x}^n, \theta_1) > c f_{X^n}(\boldsymbol{x}^n, \theta_0);$$
- 反之，若 $\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] < \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]$，则样本点 $\boldsymbol{x}^n$ 处于检验 $T(\boldsymbol{X}^n)$ 的接受域 $\mathbb{A}_n(c)$ 中，因而有
$$f_{X^n}(\boldsymbol{x}^n, \theta_1) \le c f_{X^n}(\boldsymbol{x}^n, \theta_0)\text{。}$$
- 上述任何一种情况发生，都有
$$\{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\}[f_{X^n}(\boldsymbol{x}^n, \theta_1) - c f_{X^n}(\boldsymbol{x}^n, \theta_0)] \ge 0$$
- 故
$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\}[f_{X^n}(\boldsymbol{x}^n, \theta_1) - c f_{X^n}(\boldsymbol{x}^n, \theta_0)] d\boldsymbol{x}^n \ge 0$$

证明 (Cont.):

- 由此可推

$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_1) d\boldsymbol{x}^n$$
$$\geq c \int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n$$

- 因

$$\int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n \leq \alpha,$$
$$\int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n = \alpha, \text{且} c \geq 0,$$

得

$$c \int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n \geq 0$$

- 从而 $\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{1n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_1) d\boldsymbol{x}^n \geq 0$
- 因此 $P[\boldsymbol{X}^n \in \mathbb{C}_n(c) \mid \mathbb{H}_A] \geq P(\boldsymbol{X}^n \in \mathbb{C}_{1n} \mid \mathbb{H}_A)$
- 即在 $\mathbb{H}_A$ 下，检验 $T(\boldsymbol{X}^n)$ 较 $T_1(\boldsymbol{X}^n)$ 的功效更大。

证明 (Cont.):

- **(2)** 假设 $T(\boldsymbol{X}^n)$ 是满足条件 (a)、(b) 以及 $c > 0$ 的一个检验。
- **(i) 首先证明**，任何水平为 α 的一致最大功效检验（记作 $T_2(\boldsymbol{X}^n)$) 是尺度为 α 的检验，即满足条件 (b)。
- 假设 $T_2(\boldsymbol{X}^n)$ 不是尺度为 α 的检验，则

$$\int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n < \alpha。$$

- 因为给定的检验 $T(\boldsymbol{X}^n)$ 满足条件 (b) 的假设，故有

$$\int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n = \alpha。$$

- 则

$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n > 0$$

证明 (Cont.):

- 注意:
  - ✓ 如果 $\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}] > 0$，则样本点 $\boldsymbol{x}^n$ 处于检验 $T(\boldsymbol{X}^n)$ 的拒绝域 $\mathbb{C}_n(c)$ 中，并有
  $$f_{X^n}(\boldsymbol{x}^n, \theta_1) > c f_{X^n}(\boldsymbol{x}^n, \theta_0);$$
  - ✓ 如果 $\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}] < 0$，则 $\boldsymbol{x}^n$ 处于检验 $T(\boldsymbol{X}^n)$ 的接受域 $\mathbb{A}_n(c)$ 中，并有
  $$f_{X^n}(\boldsymbol{x}^n, \theta_1) \le c f_{X^n}(\boldsymbol{x}^n, \theta_0)。$$
- 因此

$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}]\}[f_{X^n}(\boldsymbol{x}^n, \theta_1) - c f_{X^n}(\boldsymbol{x}^n, \theta_0)] d\boldsymbol{x}^n \ge 0$$

证明 (Cont.):

- 给定 $c > 0$，进一步可得

$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_1) d\boldsymbol{x}^n$$
$$\geq c \int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_{2n}]\} f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{X}^n$$
$$> 0$$

- 这表明

$$P[[\boldsymbol{X}^n \in \mathbb{C}_n(c)]|\mathbb{H}_A] > P(\boldsymbol{X}^n \in \mathbb{C}_{2n}|\mathbb{H}_A),$$

即检验 $T_2(\boldsymbol{X}^n)$ 并非一致最大功效检验，假设与结论矛盾。

- 因此， $T_2(\boldsymbol{X}^n)$ 必满足条件 (b)。

证明 (Cont.)：

- **(ii) 现在证明**，除了满足 $P(\boldsymbol{X}^n \in \mathbb{A}|\mathbb{H}_0) = P(\boldsymbol{X}^n \in \mathbb{A}|\mathbb{H}_A) = 0$ 所构成的零概率的集合 $\mathbb{A}$ 外，任一水平为 α 的一致最大功效检验必满足条件 (a) 与 (b)。
- 假设 $T^*(\boldsymbol{X}^n)$ 是任意一个水平为 α 的一致最大功效检验，其拒绝域为 $\mathbb{C}_n^*$。
- 因为任一水平为 α 的最大功效检验均是尺度为 α 的检验，故有

$$\int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n^*] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n = \alpha = \int_{\mathbb{R}^n} \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] f_{X^n}(\boldsymbol{x}^n, \theta_0) d\boldsymbol{x}^n$$

- 同样，因为 $T(\boldsymbol{X}^n)$ 和 $T^*(\boldsymbol{X}^n)$ 均为最大功效检验，故二者在 $\mathbb{H}_A$ 下功效相等：

$$\int_{\mathbb{R}^n} \{\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n^*]\} f_{X^n}(\boldsymbol{x}^n, \theta_1) d\boldsymbol{x}^n = 0$$

证明 (Cont.):

- 因此

$$\int_{\mathbb{R}^n}\{\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n(c)]-\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n^*]\}[f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_1)-cf_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_0)]d\boldsymbol{x}^n=0$$

  - ✓ 同理，若 $\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n(c)]-\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n^*]>0$，则 $\boldsymbol{x}^n$ 必处于检验 $T(\boldsymbol{X}^n)$ 的拒绝域 $\mathbb{C}_n(c)$ 中，因而有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_1)-cf_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_0)>0\,;$$

  - ✓ 反之，若 $\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n(c)]-\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n^*]<0$，因而 $\boldsymbol{x}^n$ 必处于检验 $T(\boldsymbol{X}^n)$ 的接受域 $\mathbb{A}_n(c)$ 中，并有

$$f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_1)-cf_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_0)\leq 0。$$

- 在这两种情形下，均有乘积

$$\{\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n(c)]-\mathbf{1}[\boldsymbol{x}^n\in\mathbb{C}_n^*]\}\,[f_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_1)-cf_{\boldsymbol{X}^n}(\boldsymbol{x}^n,\theta_0)]\geq 0。$$

证明 (Cont.)：

- 由于该非负的乘积的积分为 0，故除了一个零概率集合 $\mathbb{A}$ 外，对于 $\boldsymbol{X}^n$ 的样本空间上的所有样本点 $\boldsymbol{x}^n$，必有

$$\mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n(c)] - \mathbf{1}[\boldsymbol{x}^n \in \mathbb{C}_n^*] = 0。$$

- 因此，除了一个零概率集合 $\mathbb{A}$ 外， $\mathbb{C}_n^*$ 和 $\mathbb{C}_n(c)$ 是等价的，即 $T^*(\boldsymbol{X}^n)$ 满足条件 (a)。

证毕。

推论 9.1 [似然比检验和充分统计量 (Likelihood Ratio Test and Sufficient Statistic)]

- 假设 $T(\boldsymbol{X}^n)$ 是 θ 的充分统计量，$g(t,\theta_i)$ 是对应 θ_i $(i=0,1)$ 时 $T(\boldsymbol{X}^n)$ 的 PMF/PDF。
- 若给定 $c \geq 0$，定义一个检验的**拒绝域**和**接受域**分别为

$$\mathbb{C}_n(c)=\left\{t:\frac{g(t,\theta_1)}{g(t,\theta_0)}>c\right\}$$

和

$$\mathbb{A}_n(c)=\left\{t:\frac{g(t,\theta_1)}{g(t,\theta_0)}\leq c\right\}$$

其中 $P\left[T(\boldsymbol{X}^n)\in\mathbb{C}_n(c)|\mathbb{H}_0\right]=\alpha$。
- 则任何基于充分统计量 $T(\boldsymbol{X}^n)$ 的拒绝域为 $\mathbb{C}_n(c)$ 的检验都是关于 $\mathbb{H}_0$: $\theta=\theta_0$ 和 $\mathbb{H}_A$: $\theta=\theta_1$ 的水平为 α 的**一致最大功效检验**。

例 9.6

- 假设 $\boldsymbol{X}^n$ 为来自指数分布 $EXP(\theta)$ 的 IID 随机样本。
- 求在原假设 $\mathbb{H}_0: \theta = 1$ 和备择假设 $\mathbb{H}_A: \theta = 2$ 下，在 α 显著水平上的一致最大功效检验。

解：

- 指数分布 $EXP(\theta)$ 的 PDF 为 $f(x,\theta) = \frac{1}{\theta} e^{-x/\theta}, x \geq 0$。
- 因此，随机样本 $\boldsymbol{X}^n$ 的**似然函数**为

$$
\begin{aligned}
f_{\boldsymbol{X}^n}(\boldsymbol{x}^n, \theta) &= \prod_{i=1}^{n} f(x_i, \theta) \\
&= \frac{1}{\theta^n} e^{-n\bar{x}_n/\theta}
\end{aligned}
$$

- 其中 $\bar{x}_n = n^{-1} \sum_{i=1}^{n} x_i$。

解 (Cont.)：

- 由定理 6.10 因子分解定理知，$\bar{X}_n$ 是 θ 的充分统计量。
- 因为 IID $EXP(\theta)$ 随机变量之和 $\sum_{i=1}^{n} X_i$ 服从伽玛分布 $Gamma(n,\theta)$ (参见例 5.34)，故**样本均值** $\bar{X}_n \sim \mathrm{Gamma}\left(n,\frac{\theta}{n}\right)$，其 PDF

$$g(\bar{x}_n,\theta)=\frac{n^n}{(n-1)!\theta^n}\bar{x}_n^{n-1}e^{-n\bar{x}_n/\theta}, \bar{x}_n \geq 0$$

- 则**似然比**

$$\frac{f_{X^n}(x^n,\theta_1)}{f_{X^n}(x^n,\theta_0)}=\frac{g(\bar{x}_n,\theta_1)}{g(\bar{x}_n,\theta_0)}$$

$$=\frac{\frac{1}{2^n}e^{-\frac{n}{2}\bar{x}_n}}{e^{-n\bar{x}_n}}$$

$$=\frac{1}{2^n}e^{\frac{n}{2}\bar{x}_n}$$

解 (Cont.)：

- 用如下一维拒绝域定义一个检验

$$\bar{x}_n \in \mathbb{C}_n(c), \text{若 } \frac{1}{2^n} e^{\frac{n}{2}\bar{x}_n} > c$$

或等价地

$$\bar{x}_n \in \mathbb{C}_n(c), \text{若 } \bar{x}_n > 2\ln 2 + \frac{2\ln c}{n}$$

- 同时，确定常数 c 值以使得检验尺度为 α，即要求 c 满足

$$\begin{aligned} \alpha &= \int_{2\ln 2+2n^{-1}\ln c}^{\infty} g(\bar{x}_n, \theta_0) d\bar{x}_n \\ &= \int_{2\ln 2+2n^{-1}\ln c}^{\infty} \frac{n^n}{(n-1)!} \bar{x}_n^{n-1} e^{-n\bar{x}_n} d\bar{x}_n \end{aligned}$$

- 求解该非线性方程，得 $c = c(\alpha, n)$，这是 α 和 n 的函数，但它没有解析解。由内曼-皮尔逊引理与推论 9.1 可知，上述检验是尺度为 α 的**一致最大功效检验。**

目　录

- 本章后续部分，将考虑如何构建统计检验量以检验如下参数假设

$$\mathbb{H}_0: g(\theta) = \mathbf{0}$$

和

$$\mathbb{H}_A: g(\theta) \neq \mathbf{0}$$

其中

✓ $g: \mathbb{R}^p \to \mathbb{R}^J$ 是 $p \times 1$ 维参数向量 θ 的连续可导 $J \times 1$ 维向量值函数，

✓ 整数 J 是参数向量 θ 所受的约束个数。

- 假设 $J \leq p$，即约束个数不超过参数个数。

- $g(\theta)$ 的一个例子为线性向量值函数：$g(\theta) = R\theta - r$，其中
  - ✓ R 为 $J \times p$ 已知常数矩阵，
  - ✓ r 是 $J \times 1$ 已知常数向量。
- 此时原假设

$$\mathbb{H}_0 : R\theta = r$$

对 $p \times 1$ 维参数向量 θ 附加 J 个线性约束。

- 常数矩阵 R 可视为**选择矩阵** (selection matrix)。
- 例如，若选择 R 为 $p \times p$ 单位矩阵且 r 为 $p \times 1$ 零向量，则由 $R\theta = r$ 可推出参数向量 θ 的所有 p 个元素联合为零。

- 统计推断有**三大经典检验方法**：
  - ✓ **沃尔德 (Wald) 检验**
  - ✓ **拉格朗日乘子 (Lagrange multiplier, LM) 检验**
  - ✓ **似然比 (likelihood ratio, LR) 检验**
- 本章假定 $\boldsymbol{X}^n$ 为来自总体分布 $f_X(x)=f(x,\theta_0)$ 的随机样本，其中 $\theta_0\in\Theta$ 为未知真实参数值；需要检验的参数假设

$$\mathbb{H}_0: g(\theta)=\mathbf{0}$$

和

$$\mathbb{H}_A: g(\theta)\neq\mathbf{0}$$

沃尔德检验简介

- 沃尔德 (Wald) 检验的**主要思想**：
  - ✓ 根据第八章内容，可以得到 $\hat{\theta}-\theta_0 \xrightarrow{p} 0$。
  - ✓ 由 $g(\cdot)$ 函数的连续性，可知在原假设条件成立下，$g(\hat{\theta})-g(\theta_0) \xrightarrow{p} 0$。
  - ✓ 通过原假设 $\mathbb{H}_0$ 下无约束估计量与其假设值之间的加权平方距离评估模型参数约束条件的有效性，其中权重是估计量方差的逆，表示估计量的准确性。
- 然而，推导沃尔德检验的有限样本分布十分困难，因此可以推导原假设 $\mathbb{H}_0$ 下沃尔德检验的渐近分布，再据此确定统计显著性。

首先讨论**沃尔德检验**。以下是一组**正则条件**。

- **假设 9.1** $\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{d}N(0,V)$ 其中 V 为 $p\times p$ 对称有界非奇异矩阵，未知真实参数值 θ_0 是紧参数空间 Θ 的内点。
- **假设 9.2** 当 $n\to\infty$ 时，$\hat{V}\xrightarrow{p}V$。
- **假设 9.3** $g:\mathbb{R}^p\to\mathbb{R}^J$ 是 $\theta\in\Theta$ 的连续可导函数，且 $J\times p$ 梯度矩阵 $G(\theta_0)=\frac{\partial}{\partial\theta}g(\theta_0)$ 的秩为 J，其中 $J\le p$。

- **假设** 9.1 允许估计量 $\hat{\theta}$ 为任意的收敛速度为 $\sqrt{n}$ 的一致渐近正态估计量，MLE 和 MME 估计量是其中两个例子。
- **假设** 9.2 假定 $\hat{V}$ 是 $\sqrt{n}(\hat{\theta}-\theta_0)$ 的渐近方差 V 的一致估计量，且有如下的渐近展开

$$\sqrt{n}(\hat{\theta}-\theta_0)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\psi(X_i,\theta_0)+o_p(1)$$

其中 $\boldsymbol{X}^n=(X_1,\cdots,X_n)$ 是 IID 随机序列，$p\times 1$ 函数向量 $\psi(X_i,\theta_0)$ 满足 $E[\psi(X_i,\theta_0)]=\mathbf{0}$ 和 $E\parallel\psi(X_i,\theta_0)\parallel^2<\infty$，此处期望 $E(\cdot)$ 定义在总体分布 $f_X(x)$ 上。

- 当概率模型正确设定时，有 $f_X(x) = f(x, \theta_0)$。在上述假设下，$V = E[\psi(X_i, \theta_0)\psi(X_i, \theta_0)']$。因此，可构造 V 的**一致估计量**为

$$\hat{V} = \frac{1}{n}\sum_{i=1}^{n} \psi(X_i, \hat{\theta})\psi(X_i, \hat{\theta})'$$

- 由定理 7.3 关于 IID 随机样本一致强大数定律可知，适当的正则条件可保证了当 $n \to \infty$ 时几乎处处有 $\hat{V} \to V$。

- **假设 9.3** 是关于约束函数 $g(\cdot)$ 的正则条件，其中 $J\times p$ 矩阵 $G(\theta_0)$ 的满秩条件以及 $J\le p$ 确保了 $J\times J$ 对称矩阵 $G(\theta_0)VG(\theta_0)'$ 是非奇异的。

$$G(\theta)=\begin{pmatrix} \frac{\partial g_1(\theta)}{\partial\theta_1} & \frac{\partial g_1(\theta)}{\partial\theta_2} & \cdots & \frac{\partial g_1(\theta)}{\partial\theta_p} \\ \frac{\partial g_2(\theta)}{\partial\theta_1} & \frac{\partial g_2(\theta)}{\partial\theta_2} & \cdots & \frac{\partial g_2(\theta)}{\partial\theta_p} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial g_J(\theta)}{\partial\theta_1} & \frac{\partial g_J(\theta)}{\partial\theta_2} & \cdots & \frac{\partial g_J(\theta)}{\partial\theta_p} \end{pmatrix}$$

- 为检验 $\mathbb{H}_0: g(\theta_0) = \mathbf{0}$，一个自然的方法是构造一个基于统计量 $g(\hat{\theta})$ 的检验。
- 因为 $\hat{\theta}$ 是 θ_0 的一致估计量，且 $g(\cdot)$ 连续，由第七章引理 7.5 可知，当 $\hat{\theta} \xrightarrow{p} \theta_0$ 时，有 $g(\hat{\theta}) \xrightarrow{p} g(\theta_0)$。从而 $g(\hat{\theta})$ 在假设 $\mathbb{H}_0$ 下将趋近零，而在 $\mathbb{H}_A$ 之下将收敛于非零极限。
- 故可通过验证 $g(\hat{\theta})$ 是否接近零来检验 $\mathbb{H}_0$。
  - ✓ 若 $g(\hat{\theta})$ 接近零，则 $\mathbb{H}_0$ 成立。
  - ✓ 否则，$\mathbb{H}_A$ 成立。
- 现在，应用第七章的渐近理论来推导基于 $g(\hat{\theta})$ 的检验统计量的渐近分布。

- 根据**中值定理** (Bartle, 1976)，有
$$g(\hat{\theta}) = g(\theta_0) + G(\bar{\theta})\big(\hat{\theta} - \theta_0\big)$$
其中，对 $\lambda \in [0,1]$，有 $\bar{\theta} = \lambda\hat{\theta} + (1-\lambda)\theta_0$，且**梯度函数**
$$G(\theta) = \frac{dg(\theta)}{d\theta}$$
是 $J \times p$ 矩阵，其第 i 行对应 $g(\theta)$ 的第 i 个元素对参数向量 θ 每个元素的导数。

- 由于
$$\|\bar{\theta} - \theta_0\| = \|\lambda(\hat{\theta} - \theta_0)\| \le \|\hat{\theta} - \theta_0\| \xrightarrow{p} 0,$$
且 $G(\cdot)$ 连续，由连续性引理 7.5，当 $n \to \infty$ 时，有 $G(\bar{\theta}) \xrightarrow{p} G(\theta_0)$。

- 又由**渐近正态性假设**，

$$\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0,V)$$

且由**斯勒茨基定理** (参考定理 7.8)，有

$$\sqrt{n}[g(\hat{\theta})-g(\theta_0)] \xrightarrow{d} N[0,G(\theta_0)VG(\theta_0)']$$

- 在 $\mathbb{H}_0: g(\theta_0)=0$ 下，

$$\sqrt{n}g(\hat{\theta}) \xrightarrow{d} N[0,G(\theta_0)VG(\theta_0)']$$

- 给定 $G(\theta_0)$ 满秩以及 V 非奇异的条件下，$J\times J$ 矩阵 $G(\theta_0)VG(\theta_0)'$ 为非奇异矩阵，二次型

$$\sqrt{n}g(\hat{\theta})'[G(\theta_0)VG(\theta_0)']^{-1}\sqrt{n}g(\hat{\theta}) \xrightarrow{d} \chi_J^2$$

- 由 $G(\cdot)$ 的连续性以及 $\hat{\theta} \xrightarrow{p} \theta_0$，有 $G(\hat{\theta}) \xrightarrow{p} G(\theta_0)$，且由假设 9.2，

$$G(\hat{\theta})\hat{V}G(\hat{\theta})' \xrightarrow{p} G(\theta_0)VG(\theta_0)'$$

- 因此，对足够大的样本容量 n，随机矩阵 $G(\hat{\theta})\hat{V}G(\hat{\theta})'$ 为非奇异，且由斯勒茨基定理（参见定理 7.8），**沃尔德检验统计量**定义为

$$W = ng(\hat{\theta})'\left[G(\hat{\theta})\hat{V}G(\hat{\theta})'\right]^{-1}g(\hat{\theta}) \xrightarrow{d} \chi_J^2$$

其中渐近分布 χ_J^2 在 $\mathbb{H}_0$ 成立时获得。

定理 9.2 [沃尔德检验 (Wald Test)]

假定假设 9.1-9.3 以及 $\mathbb{H}_0$ 成立，则当 $n \to \infty$ 时，

$$W = ng(\hat{\theta})'\left[G(\hat{\theta})\hat{V}G(\hat{\theta})'\right]^{-1}g(\hat{\theta}) \xrightarrow{d} \chi_J^2$$

- 沃尔德检验统计量 W 是 $\sqrt{n}g(\hat{\theta})$ 和 $\sqrt{n}g(\theta_0)$ 之差的二次型，其中以 $\sqrt{n}[g(\hat{\theta}) - g(\theta_0)]$ 的渐近方差估计量 $G(\hat{\theta})\hat{V}G(\hat{\theta})'$ 的逆为权重。
- 当 W 超过渐近分布 χ_J^2 的 $(1-\alpha)$ 临界值时，沃尔德检验在显著水平 α 上拒绝原假设 $\mathbb{H}_0$: $g(\theta_0) = \mathbf{0}$。

- 另一方面，在 $\mathbb{H}_A: g(\theta_0) \neq \mathbf{0}$ 下，有

$$g(\hat{\theta}) \xrightarrow{p} g(\theta_0) \neq \mathbf{0}, G(\hat{\theta}) \xrightarrow{p} G(\theta_0), \text{且 } \hat{V} \xrightarrow{p} V$$

- 因此

$$\frac{W}{n} \xrightarrow{p} g(\theta_0)'[G(\theta_0)VG(\theta_0)']^{-1}g(\theta_0) > 0$$

- 换言之，沃尔德统计量 W 依概率以速度 n 发散至正无穷，从而确保在备择假设 $\mathbb{H}_A$ 下，当 $n \to \infty$ 时，沃尔德检验统计量 W 在任意给定的显著水平 α 上的渐近功效趋于 1。
- 这说明沃尔德检验是一致检验，因其可检测出 $\mathbb{H}_0$ 的所有备择假设。

- 需要注意，不同的估计量 $\sqrt{n}\hat{\theta}$ 有不同的渐近方差 V，因而需要不同的渐近方差估计量 $\hat{V}$。
- 现在，考察一个特例，即当 $\hat{\theta}$ 是 MLE 估计量时。此时，$V = -H^{-1}(\theta_0)$，因此可用渐近方差估计量 $\hat{V} = [-\widehat{H}(\hat{\theta})]^{-1}$，其中样本黑塞矩阵

$$\widehat{H}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \ln f(X_i,\theta)}{\partial\theta\partial\theta'}$$

- 所构造的沃尔德检验统计量如下

$$W = ng(\hat{\theta})'\left[-G(\hat{\theta})\widehat{H}(\hat{\theta})^{-1}G(\hat{\theta})'\right]^{-1}g(\hat{\theta})$$

- 若再加上第八章第三节的正则条件，可证当 $n\to\infty$ 时，几乎处处有 $\widehat{H}(\hat{\theta})\to H(\theta_0)$，从而在 $\mathbb{H}_0$ 之下 $W \xrightarrow{d} \chi_J^2$。

目　录

- 假设 X^n 为来自总体分布 $f(x,\theta_0)$ 的 IID 随机样本，其中 θ_0 是 Θ 中的未知真实参数值。
- 定义**标准化的对数似然函数**

$$\hat{l}(\theta)=\frac{1}{n}\sum_{i=1}^{n}\ln f(X_i,\theta)$$

- 现在考察在**约束条件** $g(\theta)=\mathbf{0}$ 下，如何求解如下**有约束的极大似然估计量**

$$\tilde{\theta}=\arg\max_{\theta\in\Theta}\hat{l}(\theta)$$

- 定义**拉格朗日函数**

$$L(\theta,\lambda)=\hat{l}(\theta)+\lambda' g(\theta)$$

其中 λ 为 $J\times 1$ 拉格朗日乘子向量。

- 令 $\tilde{\lambda}$ 为对应的 λ 的估计值。则**一阶条件**为

$$\frac{\partial L(\tilde{\theta},\tilde{\lambda})}{\partial \theta}=\frac{\partial \hat{l}(\tilde{\theta})}{\partial \theta}+G(\tilde{\theta})'\tilde{\lambda}=\mathbf{0}$$

$$\frac{\partial L(\tilde{\theta},\tilde{\lambda})}{\partial \lambda}=g(\tilde{\theta})=\mathbf{0}$$

- 由**中值定理**，有

$$G(\tilde{\theta})'\tilde{\lambda}=-\frac{d\hat{l}(\tilde{\theta})}{d\theta}=-\frac{d\hat{l}(\theta_0)}{d\theta}-\frac{d^2\hat{l}(\bar{\theta}_a)}{d\theta d\theta'}(\tilde{\theta}-\theta_0)$$

其中，存在 $a\in[0,1]$ 有 $\bar{\theta}_a=a\tilde{\theta}+(1-a)\theta_0$。

- 注意到

$$\frac{d^2\hat{l}(\theta)}{d\theta d\theta'}=\widehat{H}(\theta)$$

为样本**黑塞矩阵**。

- 给定第八章第三节的正则条件，上节已证当 $n \to \infty$ 时，对 θ_0 的任意强一致估计量 $\hat{\theta}$，几乎处处有

$$\hat{H}(\hat{\theta}) \to H(\theta_0)$$

- 因 $H(\theta_0)$ 非奇异，故当 $n \to \infty$ 时，几乎处处有

$$\hat{H}^{-1}(\hat{\theta}) \to H^{-1}(\theta_0)$$

且对充分大的 n，逆矩阵 $\hat{H}^{-1}(\hat{\theta})$ 存在，于是有

$$\hat{H}\left(\bar{\theta}_a\right)^{-1} G(\tilde{\theta})' \tilde{\lambda} = -\hat{H}\left(\bar{\theta}_a\right)^{-1} \frac{d\hat{l}(\theta_0)}{d\theta} - \left(\tilde{\theta} - \theta_0\right) \qquad (9.1)$$

- 另一方面，由**中值定理**，得
$$\mathbf{0}=g(\tilde{\theta})=g(\theta_0)+G(\bar{\theta}_b)(\tilde{\theta}-\theta_0)$$
- 其中，存在 $b\in[0,1]$，有 $\bar{\theta}_b=b\tilde{\theta}+(1-b)\theta_0$，故在 $\mathbb{H}_0\colon g(\theta_0)=\mathbf{0}$ 下有
$$G(\bar{\theta}_b)(\tilde{\theta}-\theta_0)=\mathbf{0} \tag{9.2}$$
- 现在，用 $G(\bar{\theta}_b)$ 乘以等式 (9.1)，并应用等式 (9.2)，得

$$\begin{aligned}G(\bar{\theta}_b)\widehat{H}(\bar{\theta}_a)^{-1}G(\tilde{\theta})'\tilde{\lambda}&=-G(\bar{\theta}_b)\widehat{H}(\bar{\theta}_a)^{-1}\frac{d\hat{l}(\theta_0)}{d\theta}-G(\bar{\theta}_b)(\tilde{\theta}-\theta_0)\\&=-G(\bar{\theta}_b)\widehat{H}(\bar{\theta}_a)^{-1}\frac{d\hat{l}(\theta_0)}{d\theta}\end{aligned}$$

- 由函数 $G(\cdot)$ 的连续性，以及当 $n\to\infty$ 时几乎处处有
  - ✓ $\|\bar{\theta}_a-\theta_0\|\le\|\tilde{\theta}-\theta_0\|\to 0$,
  - ✓ $\|\bar{\theta}_b-\theta_0\|\le\|\tilde{\theta}-\theta_0\|\to 0$,
  - ✓ $\hat{H}(\hat{\theta})\to H(\theta_0)$,
- 因此几乎处处有

$$G(\bar{\theta}_b)\hat{H}(\bar{\theta}_a)^{-1}G(\tilde{\theta})'\to G(\theta_0)H^{-1}(\theta_0)G(\theta_0)'$$

且后者非奇异。故对充分大的 n，$G(\bar{\theta}_b)\hat{H}(\bar{\theta}_a)^{-1}G(\tilde{\theta})'$ 为 $J\times J$ 非奇异矩阵。因此

$$\sqrt{n}\tilde{\lambda}=-\left[G(\bar{\theta}_b)\hat{H}(\bar{\theta}_a)^{-1}G(\tilde{\theta})'\right]^{-1}G(\bar{\theta}_b)\hat{H}(\bar{\theta}_a)^{-1}\sqrt{n}\frac{d\hat{l}(\theta_0)}{d\theta}$$

$$=-\hat{A}\sqrt{n}\frac{d\hat{l}(\theta_0)}{d\theta}$$

- 由定理 7.6 关于 IID 随机序列的**中心极限定理**，有

$$\sqrt{n}\frac{d\hat{l}(\theta_0)}{d\theta} = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \ln f(X_i,\theta_0)}{\partial\theta} \xrightarrow{d} N[0, I(\theta_0)]$$

其中 $I(\theta_0)$ 是 $\theta = \theta_0$ 处的费雪信息矩阵（Fisher’s information matrix)，参见引理 8.4 的定义。(问题：此处是否用到概率模型正确设定条件？)

- 另一方面，当 $n \to \infty$ 时，几乎处处有

$$\hat{A} \to [G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1}G(\theta_0)H(\theta_0)^{-1} = A_0$$

- 根据**斯勒茨基定理** (参见定理 7.8)

$$\sqrt{n}\tilde{\lambda} \xrightarrow{d} N[0, A_0 I(\theta_0) A_0'] \sim N\{0, -[G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1}\}$$

- 其中，利用引理 8.4 的信息等式 $I(\theta_0) + H(\theta_0) = 0$，可简化

$$\begin{aligned} A_0 I(\theta_0) A_0' &= [G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1}G(\theta_0)H(\theta_0)^{-1}I(\theta_0) \\ &\quad \times H(\theta_0)^{-1}G(\theta_0)'[G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1} \\ &= -[G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1} \end{aligned}$$

- 因此，在原假设 $\mathbb{H}_0$ 下，有二次型

$$-n\tilde{\lambda}'G(\theta_0)H(\theta_0)^{-1}G(\theta_0)'\tilde{\lambda} \xrightarrow{d} \chi_J^2$$

定理 9.3 [拉格朗日乘子 (Lagrange Multiplier, LM) 检验]

- 假定假设 8.1–8.6、假设 9.3 以及 $\mathbb{H}_0$ 成立。
- 定义**拉格朗日乘子检验统计量**

$$LM = n\tilde{\lambda}' G(\tilde{\theta})[-\widehat{H}(\tilde{\theta})]^{-1} G(\tilde{\theta})'\tilde{\lambda}$$

则在 $\mathbb{H}_0$: $g(\theta_0) = \mathbf{0}$ 之下，当 $n \to \infty$ 时，有

$$LM \xrightarrow{d} \chi_J^2$$

- 当检验统计量 LM **大于**渐近分布 χ_J^2 的 $(1-\alpha)$ 分位点时， LM 检验将在显著水平 α 上**拒绝原假设** $\mathbb{H}_0$: $g(\theta_0) = \mathbf{0}$。

目　录

- 假设 $\boldsymbol{X}^n$ 为来自总体分布为 $f_X(x)=f(x,\theta_0)$ 的 IID 随机样本，其中真实参数值 θ_0 未知。 $\boldsymbol{X}^n$ 的**似然函数**如下

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

- 定义**似然比统计量** (the likelihood ratio statistic)

$$\widehat{\Lambda}=\frac{\max_{\theta\in\Theta} f_{\boldsymbol{X}^n}(\boldsymbol{X}^n,\theta)}{\max_{\theta\in\Theta_0} f_{\boldsymbol{X}^n}(\boldsymbol{X}^n,\theta)}$$

$$=\frac{\prod_{i=1}^{n} f\left(X_i,\hat{\theta}\right)}{\prod_{i=1}^{n} f\left(X_i,\tilde{\theta}\right)}$$

- 其中 $\hat{\theta}$ 和 $\tilde{\theta}$ 分别为**无约束**和**有约束 MLE 估计量**，即

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \hat{l}(\theta)$$

$$\tilde{\theta} = \arg\max_{\theta \in \Theta_0} \hat{l}(\theta)$$

其中

$$\hat{l}(\theta) = \frac{1}{n}\sum_{i=1}^{n} \ln f(X_i, \theta)$$

是对数似然函数的**样本均值**，Θ_0 是满足约束条件 $g(\theta) = \mathbf{0}$ 的参数空间 Θ，即 $\Theta_0 = \{\theta \in \Theta : g(\theta) = \mathbf{0}\}$。

- 假定原假设 $\mathbb{H}_0: g(\theta_0) = \mathbf{0}$ 成立。则无约束和有约束 MLE 估计量 $\hat{\theta}$ 和 $\tilde{\theta}$ 均为 θ_0 的一致估计。施加约束不会导致对数似然函数的大幅下降，故似然比 $\hat{\Lambda}$ 将接近 1。换言之，若观测数据支持约束条件，则两个极大似然值的偏离不会显著超过抽样误差。
- 若 $\mathbb{H}_0$ 不成立，则无约束 MLE $\hat{\theta}$ 是 θ_0 的一致估计，但有约束 MLE $\tilde{\theta}$ 则不是。无约束似然值将会显著高于有约束似然值，赋予检验拒绝 $\mathbb{H}_A$ 的功效。此时似然比 $\hat{\Lambda}$ 将显著大于 1。因此，可通过比较 $\hat{\Lambda}$ 是否显著大于 1 或等价地比较 $\ln\hat{\Lambda}$ 是否显著大于 0 来检验原假设 $\mathbb{H}_0$。

- 正式定义**似然比检验统计量**

$$LR = 2\ln\hat{\Lambda} = 2n[\hat{l}(\hat{\theta}) - \hat{l}(\tilde{\theta})]$$

- 将 $\hat{l}(\tilde{\theta})$ 在**无约束 MLE** 估计量 $\hat{\theta}$ 处作**二阶泰勒级数展开**，得

$$LR = 2n\left\{\hat{l}(\hat{\theta}) - \left[\hat{l}(\hat{\theta}) + \frac{d\hat{l}(\hat{\theta})}{d\theta}(\tilde{\theta} - \hat{\theta}) + \frac{1}{2}(\tilde{\theta} - \hat{\theta})' \frac{d^2\hat{l}(\bar{\theta}_a)}{d\theta d\theta'}(\tilde{\theta} - \hat{\theta})\right]\right\}$$

$$= \sqrt{n}(\tilde{\theta} - \hat{\theta})'\left[-\hat{H}(\bar{\theta}_a)\right]\sqrt{n}(\tilde{\theta} - \hat{\theta})$$

其中

✓ 存在 $a \in [0,1]$，有 $\bar{\theta}_a = a\tilde{\theta} + (1-a)\hat{\theta}$，

✓ $\frac{d}{d\theta}\hat{l}(\hat{\theta}) = \mathbf{0}$ 是无约束 MLE $\hat{\theta}$ 的**一阶条件**，

✓ $\hat{H}(\theta) = \frac{d^2\hat{l}(\theta)}{d\theta d\theta'}$ 是样本**黑塞矩阵**。

- 以下，在**有约束 MLE** 估计量 $\tilde{\theta}$ 处对 $\frac{d}{d\theta}\hat{l}(\hat{\theta})$ 作**泰勒展开**，得

$$\mathbf{0}=\frac{d\hat{l}(\hat{\theta})}{d\theta}=\frac{d\hat{l}(\tilde{\theta})}{d\theta}+\widehat{H}\left(\bar{\theta}_b\right)(\hat{\theta}-\tilde{\theta})$$

- 其中，存在 $b\in[0,1]$，$\bar{\theta}_b=b\tilde{\theta}+(1-b)\hat{\theta}$。该结果与有约束 MLE 估计量的一阶条件 $G(\tilde{\theta})'\tilde{\lambda}=-\frac{d}{d\theta}\hat{l}(\tilde{\theta})$ 相结合，可推出

$$\begin{aligned}\sqrt{n}(\hat{\theta}-\tilde{\theta})&=-\widehat{H}\left(\bar{\theta}_b\right)^{-1}\sqrt{n}\frac{d\hat{l}(\tilde{\theta})}{d\theta}\\&=\widehat{H}\left(\bar{\theta}_b\right)^{-1}G(\tilde{\theta})'\sqrt{n}\tilde{\lambda}\end{aligned}$$

- 这一关系式为拉格朗日乘子估计量 $\tilde{\lambda}$ 提供了另一种解释，即其测度了无约束和有约束 MLE 估计量 $\hat{\theta}$ 和 $\tilde{\theta}$ 之间的差异。
- 故而有

$$LR=-\sqrt{n}\tilde{\lambda}'G(\tilde{\theta})\widehat{H}\left(\bar{\theta}_b\right)^{-1}\widehat{H}\left(\bar{\theta}_a\right)\widehat{H}\left(\bar{\theta}_b\right)^{-1}G(\tilde{\theta})'\sqrt{n}\tilde{\lambda}$$

- 因为

$$LR - LM = -\sqrt{n}\tilde{\lambda}'\left[G(\tilde{\theta})\widehat{H}(\bar{\theta}_b)^{-1}\widehat{H}(\bar{\theta}_a)\widehat{H}(\bar{\theta}_b)^{-1}G(\tilde{\theta})' - G(\tilde{\theta})\widehat{H}(\tilde{\theta})^{-1}G(\tilde{\theta})'\right]\sqrt{n}\tilde{\lambda}$$
$$= O_p(1)o_p(1)O_p(1)$$
$$= o_p(1)$$

其中，由引理 7.11 以及

$$\sqrt{n}\tilde{\lambda} \xrightarrow{d} N\{\mathbf{0}, -[G(\theta_0)H(\theta_0)^{-1}G(\theta_0)']^{-1}\}$$

可得 $\sqrt{n}\tilde{\lambda} = O_p(1)$，且

$$G(\tilde{\theta})\widehat{H}(\bar{\theta}_b)^{-1}\widehat{H}(\bar{\theta}_a)\widehat{H}(\bar{\theta}_b)^{-1}G(\tilde{\theta})' - G(\tilde{\theta})\widehat{H}(\tilde{\theta})^{-1}G(\tilde{\theta})'$$
$$\xrightarrow{p} G(\theta_0)H(\theta_0)^{-1}H(\theta_0)H(\theta_0)^{-1}G(\theta_0)' - G(\theta_0)H(\theta_0)^{-1}G(\theta_0)' = \mathbf{0}$$

- 因此在原假设 $\mathbb{H}_0$ 下两个统计量 LR 和 LM 渐近等价。

- 据**渐近等价性引理** (参见引理 7.10) 以及在 $\mathbb{H}_0$ 下 $LM \xrightarrow{d} \chi_J^2$ (参见定理 9.3)，可得如下结论。

定理 9.4 [似然比 (Likelihood Ratio, LR) 检验]

若假设 8.1-8.6 、假设 9.3 以及 $\mathbb{H}_0: g(\theta_0) = \mathbf{0}$ 成立，则在原假设 $\mathbb{H}_0$ 下，当 $n \to \infty$ 时，似然比检验统计量

$$LR \xrightarrow{d} \chi_J^2$$

- 似然比检验、沃尔德检验与拉格朗日乘子检验是假设检验的三大经典方法。
- 还可证明，在原假设 $\mathbb{H}_0$ 下，基于 MLE $\hat{\theta}$ 的沃尔德检验统计量 W 和拉格朗日乘子检验统计量 LM 也渐近等价。
- 在模型误设时，依然可以构造服从渐近 χ^2 分布的稳健沃尔德检验统计量和稳健 LM 检验统计量，但却不能构造稳健 LR 检验统计量。

目　录

- 现在举两个简单例子说明如何计算：
  - ✓ 基于 MLE 的沃尔德检验统计量 W
  - ✓ 拉格朗日乘子检验统计量 LM
  - ✓ 似然比检验统计量 LR
- 第一个例子是来自伯努利分布 Bernoulli(θ) 的 IID 随机样本，第二个例子是来自正态分布 $N(\mu, \sigma^2)$ 的 IID 随机样本。

9.6.1 伯努利分布下的假设检验

- 假设 $\boldsymbol{X}^n$ 为来自伯努利分布 Bernoulli(θ) 的 IID 随机样本，其中伯努利随机变量有两个可能的取值：

$$X_i = \begin{cases} 1, & \text{概率为 } \theta \\ 0, & \text{概率为 } 1-\theta \end{cases}$$

- 参数 $\theta \in (0,1)$ 未知。现考虑检验

$$\mathbb{H}_0: \theta = \theta_0$$

和

$$\mathbb{H}_A: \theta \neq \theta_0$$

其中 θ_0 为已知常数。因此，有 $g(\theta) = \theta - \theta_0$，且梯度即导数为

$$G(\theta) = \frac{dg(\theta)}{d\theta} = 1$$

- 因为总体 PMF $f(x,\theta)=\theta^x(1-\theta)^{1-x}$，$x=0,1$，则 IID 随机样本 $\boldsymbol{X}^n$ 的**对数似然函数**为

$$\begin{aligned}\ln\hat{L}(\theta\mid\boldsymbol{X}^n)&=\sum_{i=1}^{n}\ln f(X_i,\theta)\\&=n\bar{X}_n\ln\theta+n(1-\bar{X}_n)\ln(1-\theta)\end{aligned}$$

- 其中样本均值 $\bar{X}_n$ 是 θ 的充分统计量。MLE 的**一阶条件**为

$$\frac{\partial\ln\hat{L}(\theta\mid\boldsymbol{X}^n)}{\partial\theta}=\frac{n\bar{X}_n}{\hat{\theta}}-\frac{n-n\bar{X}_n}{1-\hat{\theta}}=0$$

- 因此，MLE $\hat{\theta}=\bar{X}_n$。

- 首先考察沃尔德检验。回顾引理 8.4 定义的**黑塞矩阵**

$$H(\theta) = E_\theta\left[\frac{\partial^2 \ln f(X_i,\theta)}{\partial\theta^2}\right]$$

- 因为

$$\frac{\partial^2 \ln f(X_i,\theta)}{\partial\theta^2} = -\frac{X_i}{\theta^2} - \frac{1-X_i}{(1-\theta)^2}$$

- **样本黑塞矩阵**

$$\begin{aligned}\widehat{H}(\theta) &= n^{-1}\sum_{i=1}^{n}\frac{\partial^2 \ln f(X_i,\theta)}{\partial\theta^2} \\ &= -\frac{\sum_{i=1}^{n} X_i}{n\theta^2} - \frac{\sum_{i=1}^{n}(1-X_i)}{n(1-\theta)^2} \\ &= -\frac{\bar{X}_n}{\theta^2} - \frac{1-\bar{X}_n}{(1-\theta)^2}\end{aligned}$$

- 因此，有

$$\hat{H}(\hat{\theta}) = -\frac{1}{\bar{X}_n(1-\bar{X}_n)}$$

- 在原假设 $\mathbb{H}_0: \theta = \theta_0$ 下，当 $n \to \infty$ 时，**沃尔德检验统计量**为

$$\begin{aligned} W &= n g(\hat{\theta})' \left[-G(\hat{\theta})\hat{H}^{-1}(\hat{\theta})G(\hat{\theta})'\right]^{-1} g(\hat{\theta}) \\ &= \frac{n\left(\hat{\theta}-\theta_0\right)^2}{\bar{X}_n(1-\bar{X}_n)} \\ &= \frac{n(\bar{X}_n-\theta_0)^2}{\bar{X}_n(1-\bar{X}_n)} \\ &\xrightarrow{d} \chi_1^2 \end{aligned}$$

- 这里渐近分布 χ_1^2 由 $\sqrt{n}(\bar{X}_n-\theta_0) \xrightarrow{d} N\left(0,\sigma^2\right)$ 和斯勒茨基定理 (定理 7.8) 而得。

- 下面考察拉格朗日乘子检验。定义**拉格朗日函数**

$$L(\theta,\lambda)=\hat{l}(\theta)+\lambda' g(\theta)=\hat{l}(\theta)+\lambda(\theta-\theta_0)$$

其中**标准化对数似然函数**

$$\hat{l}(\theta)=\bar{X}_n\ln\theta+(1-\bar{X}_n)\ln(1-\theta)$$

- **有约束** MLE 估计量的**一阶条件**为

$$\frac{\partial L(\tilde{\theta},\tilde{\lambda})}{\partial\theta}=\frac{\partial\hat{l}(\tilde{\theta})}{\partial\theta}+\tilde{\lambda}=0$$

$$\frac{\partial L(\tilde{\theta},\tilde{\lambda})}{\partial\lambda}=g(\tilde{\theta})=\tilde{\theta}-\theta_0=0$$

- 从而 $\tilde{\theta}=\theta_0$，且

$$\begin{aligned}\tilde{\lambda} &= -\frac{\partial \hat{l}(\tilde{\theta})}{\partial \theta} \\ &= -\frac{\bar{X}_n}{\tilde{\theta}}+\frac{1-\bar{X}_n}{1-\tilde{\theta}} \\ &= -\frac{\bar{X}_n-\tilde{\theta}}{\tilde{\theta}(1-\tilde{\theta})} \\ &= -\frac{\bar{X}_n-\theta_0}{\theta_0(1-\theta_0)}\end{aligned}$$

- 这表明 $\tilde{\lambda}$ 测度了无约束 MLE $\hat{\theta}=\bar{X}_n$ 和有约束MLE $\tilde{\theta}=\theta_0$ 之间的差异。

- 同时，**样本黑塞矩阵**

$$\widehat{H}(\tilde{\theta}) = -\frac{\bar{X}_n}{\theta_0^2} - \frac{1-\bar{X}_n}{(1-\theta_0)^2}$$
$$= -\frac{\bar{X}_n(1-\theta_0)^2 + (1-\bar{X}_n)\theta_0^2}{\theta_0^2(1-\theta_0)^2}$$

- 因此，有

$$LM = -n\tilde{\lambda}'G(\tilde{\theta})\widehat{H}(\tilde{\theta})^{-1}G(\tilde{\theta})'\tilde{\lambda}$$
$$= n\left[-\frac{\bar{X}_n-\theta_0}{\theta_0(1-\theta_0)}\right]^2\left[\frac{\bar{X}_n(1-\theta_0)^2 + (1-\bar{X}_n)\theta_0^2}{\theta_0^2(1-\theta_0)^2}\right]^{-1}$$
$$= \frac{n(\bar{X}_n-\theta_0)^2}{\bar{X}_n(1-\theta_0)^2 + (1-\bar{X}_n)\theta_0^2}$$

- 最后，**似然比检验统计量**

$$\begin{aligned} LR &= 2n[\hat{l}(\hat{\theta}) - \hat{l}(\tilde{\theta})] \\ &= 2n\left[\bar{X}_n \ln\left(\frac{\bar{X}_n}{\theta_0}\right) + (1-\bar{X}_n)\ln\left(\frac{1-\bar{X}_n}{1-\theta_0}\right)\right] \end{aligned}$$

9.6.2 正态分布下的假设检验

- 假设 $\boldsymbol{X}^n$ 为来自正态分布 $N(\mu,\sigma^2)$ 的 IID 随机样本，其中参数 $\theta=(\mu,\sigma^2)$ 未知。
- 考虑检验如下假设

$$\mathbb{H}_0: \mu=\mu_0 \text{ 和 } \mathbb{H}_A: \mu \neq \mu_0$$

其中 μ_0 为已知常数。

- 这等价于选择检验函数

$$g(\theta)=\mu-\mu_0$$

- 从而

$$G(\theta)=\frac{dg(\theta)}{d\theta}=(1,\ 0)$$

是一个二维行向量。

- 因为总体分布 $N(\mu,\sigma^2)$ 的 PDF 为

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

- 随机样本 X^n 的**标准化对数似然函数**是

$$\hat{l}(\theta)=-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\sigma^2\right)-\frac{1}{2\sigma^2}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$$

- 对**无约束的 MLE**，例 8.4 已求得

$$\hat{\mu}=\bar{X}_n$$

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

- 同时，**样本黑塞矩阵**

$$\hat{H}(\theta)=\begin{bmatrix} -\frac{1}{\sigma^2} & -\frac{1}{\sigma^4}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) \\ -\frac{1}{\sigma^4}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu) & \frac{1}{2\sigma^4}-\frac{1}{\sigma^6}\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2 \end{bmatrix}$$

- 当 $\theta=\hat{\theta}$ 时，有

$$\hat{H}(\hat{\theta})=\begin{bmatrix} -\frac{1}{\hat{\sigma}^2} & 0 \\ 0 & -\frac{1}{2\hat{\sigma}^4} \end{bmatrix}$$

- 从而在原假设 $\mathbb{H}_0: \mu = \mu_0$ 下，**沃尔德检验统计量**为

$$
\begin{aligned}
W &= -ng(\hat{\theta})'\left[G(\hat{\theta})\hat{H}(\hat{\theta})^{-1}G(\hat{\theta})'\right]^{-1}g(\hat{\theta}) \\
&= \frac{n(\bar{X}_n - \mu_0)^2}{\hat{\sigma}^2} \\
&\xrightarrow{d} \chi_1^2
\end{aligned}
$$

- 这里渐近分布 χ_1^2 由 $\sqrt{n}(\bar{X}_n - \mu)/\sigma \sim N(0,1)$ 和斯勒茨基定理 (定理 7.8) 而得。

构建拉格朗日乘子检验统计量

- 考察**有约束 MLE 估计量**

$$\tilde{\theta} = \underset{\theta \in \Theta}{\operatorname{argmax}} \hat{l}(\theta)$$

- 此处约束条件为 $\mu = \mu_0$。定义**拉格朗日函数**

$$L(\theta, \lambda) = \hat{l}(\theta) + \lambda(\mu - \mu_0)$$

- 则**一阶条件**为

$$\frac{\partial L(\tilde{\theta}, \tilde{\lambda})}{\partial \mu} = \frac{1}{\tilde{\sigma}^2} \frac{1}{n} \sum_{i=1}^{n} (X_i - \tilde{\mu}) + \tilde{\lambda} = 0$$

$$\frac{\partial L(\tilde{\theta}, \tilde{\lambda})}{\partial \sigma^2} = -\frac{1}{2\tilde{\sigma}^2} + \frac{1}{2\tilde{\sigma}^4} \frac{1}{n} \sum_{i=1}^{n} (X_i - \tilde{\mu})^2 = 0$$

$$\frac{\partial L(\tilde{\theta}, \tilde{\lambda})}{\partial \lambda} = \tilde{\mu} - \mu_0 = 0$$

- **求解一阶条件**，得

$$\tilde{\mu} = \mu_0$$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \mu_0)^2$$

$$\tilde{\lambda} = -\frac{1}{\tilde{\sigma}^2}(\bar{X}_n - \mu_0)$$

且**样本黑塞矩阵**

$$\widehat{H}(\tilde{\theta}) = \begin{bmatrix} -\frac{1}{\tilde{\sigma}^2} & -\frac{1}{\tilde{\sigma}^4}(\bar{X}_n - \mu_0) \\ -\frac{1}{\tilde{\sigma}^4}(\bar{X}_n - \mu_0) & -\frac{1}{2\tilde{\sigma}^4} \end{bmatrix}$$

- 因此，**拉格朗日乘子检验统计量**

$$LM = -n\tilde{\lambda}' G(\tilde{\theta}) \widehat{H}^{-1}(\tilde{\theta}) G(\tilde{\theta})' \tilde{\lambda}$$

$$= \frac{n(\bar{X}_n - \mu_0)^2}{\tilde{\sigma}^2 - 2(\bar{X}_n - \mu_0)^2}$$

- 最后，构建**似然比检验统计量** LR。
- 因为

$$\hat{l}(\hat{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\hat{\sigma}^2\right) - \frac{1}{2}$$

$$\hat{l}(\tilde{\theta}) = -\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln\left(\tilde{\sigma}^2\right) - \frac{1}{2}$$

- 从而

$$LR = 2n[\hat{l}(\hat{\theta}) - \hat{l}(\tilde{\theta})]$$
$$= n\ln\left(\frac{\tilde{\sigma}^2}{\hat{\sigma}^2}\right)$$

- 上式表明，极大似然检验通过比较原假设和备择假设下的**方差估计量**来检验总体均值是否满足 $\mu = \mu_0$。这里似然比等于样本方差比的对数函数。
- 直观上，当 $\mu \neq \mu_0$，$\tilde{\sigma}^2 = n^{-1}\sum_{i=1}^{n}(X_i - \mu_0)^2$ 并非 σ^2 的一致估计量。
- 当 n 很大时，$\tilde{\sigma}^2$ 将大于 $\hat{\sigma}^2$，赋予 LR 拒绝原假设的功效。当 $n \to \infty$ 时，LR 检验统计量趋向无穷大。

目　录

- 假设检验是统计推断最重要的任务之一。
- 本章介绍了统计推断中假设检验的**基本概念**和**基本思想**，以及著名的**内曼-皮尔逊引理**，即基于**似然比**的检验是简单假设的**一致最大功效检验**。
- 接着，讨论了三个经典的统计检验方法：
  - ✓ 沃尔德检验
  - ✓ 拉格朗日乘子检验
  - ✓ 似然比检验
- 并证明了在参数概率模型正确设定条件下三者在原假设下渐近等价。

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！