第 10 讲*

假设检验

麻省理工学院 14.30 2006 年春季

Herman Bennett

假设检验：给定来自某一总体中的一个随机样本，这一样本是否足以否定对总体的某种猜想。（例如：参数的值）

25 定义

25.1 假设检验

一个(参数)假设是对一个或多个总体参数的一种命题陈述。[1] 该假设可以通过*假设检验*来验证。

一个假设检验包括：

1. 两个互补假设：原假设和备择假设，分别记作 H_0 和 H_1。
2. 决策规则为指出对哪些样本值不拒绝（“接受”）原假设，对哪些样本值拒绝原假设而接受备择假设。

H_0 被拒绝的样本数值的集合被称作拒绝域或判别域。判别域以外的区域被称作接受域（H_0 被接受）。

注意：这些讲义不一定是自封的。它们只是对讲座的一种补充而不是替代。

[1]非参数假设的例子应该是对随机变量X分布的一种命题陈述，例如：X～N()。

构造一个假设就能够真正完整地确定总体分布，称为简单假设；否则，称为复合假设。

25.2 假设检验的一般形式

$X_1, . . . , X_n$ 是一个总体概率密度函数为 $f(x|\theta)$ 的随机样本。定义对于参数 θ ∈ Ω，进行的假设检验为：

$$H_0\text{：}\ \theta \in \Omega$$

$$H_1\text{：}\ \theta \notin \Omega$$

其中，$\Omega_0 \cup \Omega_1 = \Omega$和$\Omega_0 \cap \Omega_1 = \varnothing$。如果随机样本 $X_1, . . . , X_n$ 落在 n 维空间 C 中，则拒绝 H_0。空间 C 是对 X 而言的一个*判别区域*，X 是包含随机样本的 n 维向量。

两个互补假设 H_0 和 H_1 通常表现为下面五种形式：

1. 简单原假设 H_0 和简单备择假设 H_1： H_0： $\theta = \theta_0$

$$H_1\text{：}\ \theta = \theta_1 \qquad \textbf{(76)}$$

2. 简单原假设 H_0 和复合双侧备择假设 H_1： H_0： $\theta = \theta_0$

$$H_1\text{：}\ \theta \neq \theta_1 \qquad (77)$$

3. 简单原假设 H_0 和简单备择假设 H_1： H_0： $\theta = \theta_0$

$$H_1\text{：}\ \theta \langle \theta_0 (\text{或}\theta \rangle \theta_0) \qquad \textbf{(78)}$$

4. 复合单侧原假设 H_0 及复合单侧备择假设 H_1： H_0： $\theta \leq \theta_0$（或$\theta \geq \theta_0$）

$$H_1\text{：}\ \theta \rangle \theta_0 (\text{或}\theta \langle \theta_0) \qquad \textbf{(79)}$$

5. 复合双侧原假设 H_0 和复合双侧备择假设 H_1： H_0： $\theta_{n_1} \leq \theta \leq \theta_{n_2}$

$$H_1\text{：}\ \theta \langle \theta_{n_1} \text{且} \theta \rangle \theta_{n_2}) \qquad \textbf{(80)}$$

25.3 假设检验的错误类型

当 H_0 为真却拒绝 H_0 时发生第一类错误，该错误发生的概率记作 α_θ，定义如下：

$$\alpha_\theta = P(\text{第一类错误}) = P(\text{拒绝 H}_0 | \theta \in \Omega) \tag{81}$$

当 H_1 为真时却接受H_0时发生第二类错误，该错误发生的概率记作 β_θ，定义如下：

$$\beta_\theta = P(\text{第二类错误}) = P(\text{接受 H}_0 | \theta \in \Omega) \tag{82}$$

- 总结：

25.3.1 显著性水平及最优检验

假设检验的显著性水平（或大小）是指发生犯第一类错误的最大概率。显著性水平记作 α。[2]

$$\alpha = \sup_{\theta \in \Omega_0} \alpha_\theta . \tag{83}$$

如果 Ω_0 是独点集，则 $\alpha = \alpha_\theta$。

对于给定的一组原假设和备择假设，在给定的显著性水平下，最优假设检验定义为对于任意θ，使 β_θ 最小化的检验。注意，很多假设检验并不存在最优检验（以后再详讲）。

[2]水平和大小之间存在一个技术差别，实践中，仅在复杂检验才细分。本课程中，我们设定两者可交换使用。

例 25.1 假设一个来自 $N(\mu, 4)$ 总体的容量为 n 的随机样本，（1）运用统计量 $\bar{X}$ 建立一个假设检验，H_0：$\mu = 0$，H_1：$\mu = 1$，而且决策规则为"当$\bar{X} \rangle k$时拒绝H_0"。这样犯第一类错误的概率为 5%。（2）计算犯第二类错误的概率。检验的置信水平是多少？（3）当 k 变大或者变少时，α 和 β 会有什么变化？如何进行两者的权衡？（4）如果样本容量 n 增加又会如何？（5）如果我们重新设定假设 H_0：$\mu = 0$，H_1：$\mu \neq 0$ 结果会有什么变化？

- 解释假设检验的结果时要注意：接受与不拒绝 H_0 的区别。

25.4 功效函数

用字母 δ 来表示假设检验（原假设，备择假设及决策规则）的特征。

假设检验 δ 的<u>功效函数</u>是指已知参数的真实值为 θ ∈ Ω，拒绝 H_0 的概率。

$$\pi_\theta(\theta|\delta)=P(\text{拒绝}H_0|\theta\in\Omega)=P(X\in C|\theta)\text{，对于所有的}\theta\in\Omega \qquad \textbf{(84)}$$

因此，

$$\pi_\theta(\theta|\delta)=\alpha_\theta\text{，如果}\theta\in\Omega_0$$

$$1-\pi_\theta(\theta|\delta)=\beta_\theta\text{，如果}\theta\in\Omega_1 \qquad \textbf{(85)}$$

例 25.2. 理想的功效函数… $a=?$，b=?

$$\pi(\theta|\delta)=\begin{cases} a & \text{如果 }\theta\in\Omega_0 \\ b & \text{如果 }\theta\in\Omega_1 \end{cases}$$

如果Ω_0 是独点集，那么$\alpha=\pi(\theta|\delta)$。

对于给定的一对原假设和备择假设，给定的显著性水平 α，最优的假设检验 δ^*是指对于所有$\theta\in\Omega_1$ 使 β (δ)最小的检验。换句话说，对于所有的$\theta\in\Omega_1$， δ^* 使功效函数最大。

例 25.3. 设一个来自总体$U[0,\theta]$容量为 n 的随机样本，θ 未知，设定下列假设检验 δ：

H_0：$3\le\theta\le4$)

H_1：$\theta\langle 3\text{或}\theta\rangle 4$)

决策规则：如果$\hat{\theta}_{MLE}\in[2.9,4.1]$，接受$H_0$，否则拒绝$H_0$。计算功效函数$\pi(\theta|\delta)$ (注： $\forall\theta$)。检验的显著性水平为多少?

25.5 $p-$值法

$p-$值是描述给定的随机样本得到数值实现，拒绝 H_0 时，显著性水平 α 的最小值。因此，无论 H_0 被拒绝或者被接受，$p-$值是后验计算法。

26 （四个）最常用的假设检验框架

26.1 似然比检验(LRT):

$$H_0: \ \theta = 0$$

$$H_1: \ \theta = 1$$

决策规则："如果 $f_1(\mathrm{x})/f_0(\mathrm{x})\rangle \mathrm{k}$，则拒绝 H_0 "　　　（**86**）

其中，根据检验的显著性水平（α_0）的大小选取常数 k>0。因此，$P(f_1(\mathrm{x})/f_0(\mathrm{x})\rangle \mathrm{k}\,|\theta)=\alpha_0$。统计量 $f_1(\mathrm{x})/f_0(\mathrm{x})_1$是根据如下表达式给定：

$$f_i(\mathrm{x})=f(x_1,x_2,...,x_n|\theta_i)=f(x_1|\theta_i)f(x_2|\theta_i)...f(x_n|\theta_i)$$（独立同分布样本）　（**87**）

- $f_1(\mathrm{x})/f_0(\mathrm{x})$的比率被称为样本的似然比统计量。

似然比检验的最优性

给定犯第一类错误的概率而使犯第二类错误的概率最小。

$$\min_{\delta} \beta; \ 给定\alpha_0$$

(α_0 是检验设定的显著性水平)

(奈曼-皮尔逊引理) 设 δ^* 是一个假设检验，其中 H_0 和 H_1 为简单假设，如果 $kf_0(\mathrm{x})\rangle f_1(\mathrm{x})$ (k⟩0)，接受 H_0；否则，接受 H_1。，此外，$kf_0(\mathrm{x})=f_1(\mathrm{x})$时 H_0 和 H_1 都被接受。那么，对于任意一个其他的假设检验 δ：

$$\beta(\delta)\langle\beta(\delta^*)\leftrightarrow\alpha(\delta)\rangle\alpha(\delta^*)$$　　（**88**）

例 26.1. 设一个来自总体服从二项分布，容量为 20 的随机样本，p 未知。设定的假设如下：

$$H_0：\ p = 0.2$$

$$H_1：\ p = 0.4$$

计算最优的检验过程 δ^*，使 $\alpha(\delta^*) = 0.05$。

- 在正态随机样本的情况下，假设检验式**(86)**意味着如下的决策规则：

—当 $\theta_1 \rangle \theta_2$ 时，如果 $\overline{x} \rangle k'$, 则“拒绝 H_0 ”。

—当 $\theta_1 \langle \theta_2$ 时，如果 $\overline{x} \langle k'$, 则“拒绝 H_0 ”。

(此结果引自 **DeGroot and Schervish** (2002) 第 465 页)

26.2 单侧检验:

$$H_0\text{：}\ \mu = \mu_0$$

$$H_1\text{：}\ \mu \rangle \mu_0 \ \text{或}\ H_1\text{：}\ \mu \langle \mu_0$$

决策规则：“如果 $\overline{x} \rangle c$, 则拒绝 H_0 或“如果 $\overline{x} \langle c$, 则拒绝 H_0 。 **（89）**

其中，根据检验的显著性水平（α_0)的大小选取常数 c。

因此，$P(\overline{X} \rangle c \mid \mu_0) = \alpha_0$ 或者 $P(\overline{X} \langle c \mid \mu_0) = \alpha_0$。

单侧检验的最优性

在这些例子中，什么意味着“最优”？如果我们使用非最优检验呢？

对这些例子中相应的最优结果的广义推导已经超出了本课程的范围[3]。然而，我们可以简单阐述以下结论：

— 设服从二项分布或正态分布的一个随机样本，式**(89)**给定的原假设和备择假设，显著性水平为 α_0。那么，对于所有 $\theta \in \Omega_1$，使 $\beta(\delta)$ 最小的最优检验 δ^* 可以由式**(89)** 检验得出。

式**(89)**的决策规则也适用于式（**78**）和式（**79**)，即使它不一定是最优检验。

3 最好参见DeGroot and Schervish (2002)第八章中 8.3 节部分。

例 26.2 设一个来自总体 $N(\mu,1)$ 的容量为 100 的随机样本。μ 未知，$\overline{x}=1.13$。

进行如下假设：

$$H_0：\mu=1$$

$$H_0：\mu\rangle 1$$

构造一个显著性水平为 0.05 的单侧假设检验。验证原假设并计算 p-值。

26.3 双侧检验：

$$H_0：\mu=\mu_0$$
$$H_1：\mu\neq\mu_0$$

决策规则为："如果 $\overline{x}\notin[c_1,c_2]$,则拒绝 H_0"　　**(90)**

根据检验的显著性水平（α_0)的大小选取常数 c_1，c_2。因此，$P(\overline{x}\notin[c_1,\ c_2]|\mu_0)=\alpha_0$。通常，假设检验是以对称的方式构造，表示为 $P(\overline{X}\langle c_1|\mu_0)=\alpha_0/2$ 和 $P(\overline{X}\rangle c_2|\mu_0)=\alpha_0/2$。

双侧检验的最优性

令人遗憾的是，这些例子中关于最优检验没有得出一个稳健的结果。对于所有 $\theta \in \Omega_1$，没有使 $\beta_\theta(\delta)$ 最小的最优检验过程 δ^*。然而，单侧假设检验最优结果的例子也可以说明式（**90**）中描述的是一个合理的假设检验决策规则。[4]

式**(90)**的决策规则也适用于式 **(77)**和式**(80)**。

例 26.3 一个蜡烛厂商声宣称他们的蜡烛平均可以持续燃烧 60 分钟。一个对此声明表示怀疑的消费者买了 40 根蜡烛，并作了试验。他发现蜡烛平均可以持续燃烧 65.22 分钟。根据搜集的数据，他计算了统计量s^2=225。此消费者能够以 99%的显著性水平，断定该厂商的声明是错误的吗？（假设样本是独立同分布）另外，再计算 p −值以及在显著性水平为 $\alpha = 0.01$ 时，拒绝 H_0 的极限 n。（假定s^2和$\bar{X}$的值保持不变）

[4] 实际上，这正是许多学者研究的问题。

26.4 广义似然比检验（GLRT):

H_0 和 H_1：任意复合假设或简单假设

决策规则表达形式为：“如果 $W \rangle K$，则拒绝 H_0” （91）

根据检验的显著性水平（α_0)的大小选取常数 $k \rangle 0$。因此，$P(W \rangle k | H_0) = \alpha_0$。统计量 W 由下式确定：

$$W = \frac{\sup_{\theta \in \Omega_1} L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n)}{\sup_{\theta \in \Omega_0} L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n)} = \frac{\sup_{\theta \in \Omega_1} f(\mathbf{x} | \theta \in \Omega_1)}{\sup_{\theta \in \Omega_0} f(\mathbf{x} | \theta \in \Omega_0)} \quad (92)$$

- 与前面的检验一样，常数 k 取决于统计量 W 的分布和 α_0。如果计算 W 的分布变得尤为艰难，那么可能会使用一个广义极大似然比等价的定义，如式（93）所示，具有一个已知的极限分布。

$$T = \frac{\sup_{\theta \in \Omega_0} L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n)}{\sup_{\theta \in \Omega} L(\theta_1, \ldots, \theta_k | x_1, \ldots, x_n)} = \frac{\sup_{\theta \in \Omega_0} f(\mathbf{x} | \theta \in \Omega_0)}{\sup_{\theta \in \Omega} f(\mathbf{x} | \theta \in \Omega)} \quad (93)$$

决策规则表达形式为：“如果 T⟨d，则拒绝 H_0”。根据检验的显著性水平（α_0)的大小选取常数 $d \rangle 0$。因此，$P(T \langle d | H_0) = \alpha_0$。$-2\ln T$ 的极限分布已知为：

$$-2lnT \overset{n \to \infty}{\sim} \chi^2_{(r)} \quad (94)$$

其中，r 等于 Ω 中的自由参数#减掉 Ω_0 中的自由参数#。如果 $-2\ln T \rangle \chi^2_{(r), \alpha}$ [5]，则拒绝 H_0。

如果能够直接计算 W 或 T 的分布，那么最好用该分布代替极限 χ^2 分布。

广义似然比的最优性

广义似然比检验是极大似然比检验的推广形式；在 H_0 或/和 H_1 为复合假设时应用该方

[5] 技术结果表明该分布是一个 $\chi_{(r)}{}^2$ 分布，自由度 $r = \dim\Omega - \dim\Omega_0$。

法。然而，广义似然比检验和极大似然比检验一样，不一定是最优的，特别的是，它要取决于具体的案例情况（关于此问题的细节超出了本课程的范围[6]）。

27 两个正态样本的假设检验

例 27.1 假设两个随机样本：

$X_i \sim N(\mu_X, \sigma_X^2)$，样本容量为$n_X$

$Y_i \sim N(\mu_Y, \sigma_Y^2)$，样本容量为$n_Y$

做如下假设：

(a) $H_0: \mu_X = \mu_Y$，$H_1: \mu_X \neq \mu_Y$　　　**(b)** $H_0: \sigma_X^2 = \sigma_Y^2$，$H_1: \sigma_X^2 \neq \sigma_Y^2$

对每种情况构造一个显著性水平为 95%的假设检验。在(**a**)部分假定已知 σ_X^2 和 σ_Y^2。

谢谢大家！

[6] 最好参见DeGroot and Schervish(2002)第 8 章。