

## 第八节 · 假设检验

司继春

上海对外经贸大学统计与信息学院

在上一节中，我们讨论了对于未知总体参数  $\theta$  的参数估计问题，包括点估计和区间估计。很多时候，我们不仅仅需要回答总体参数  $\theta$  是多少的问题，或者  $\theta$  在什么区间范围以内的问题，还要回答「是不是」的问题，比如  $\theta = \theta_0$  是不是成立？例如，在上一节中，我们发现在中国城镇住户调查的数据中，男性收入平均比女性年收入多 8222 元，那么我们是不是可以推断总体男性收入的确比女性高呢？还是仅仅因为抽样的巧合导致了我们的男性收入比女性收入高？我们可以使用**假设检验**（Hypothesis testing）的方法回答此类问题。

### 1 假设检验

为了讨论假设检验的问题，我们首先介绍「假设 (hypothesis)」的概念。在假设检验中，假设指的是关于总体参数  $\theta$  的一个命题。比如，对于不同总体，我们可能有如下假设：

1. 山东成年男性的平均身高为 175cm ( $\theta = \theta_0$ )
2. 某生产线次品率控制在 0.1% 范围内 ( $\theta \leq \theta_0$ )
3. 北方人平均身高高于南方人 ( $\theta_1 \geq \theta_2$ )

以上的例子都是关于未知总体参数的一些猜想。由于总体参数是未知的，我们只能观察到样本，因而我们不能确切的知道以上命题究竟是否成立，而只能使用样本对以上命题进行推断。

需要注意的是，这里的假设 (hypothesis) 与数学命题中的假设 (assumption) 是不同的。假设检验中的假设是我们要验证或者推翻的某个命题，而数学命题中的假设则是结论的前提条件。

假设检验中有两个互补的假设：**原假设** (null hypothesis) 和**备择假设** (alternative hypothesis)，分别用  $H_0$  和  $H_1$  来表示。如果参数的范围为  $\Theta$ ，即总体参数  $\theta \in \Theta$ ，而原假设为  $\theta \in \Theta_0$ ，那么备择假设即为原假设的补集，即  $\theta \in \Theta_0^c$ 。比如，如果  $\Theta = \mathbb{R}$ ，若原假设是  $\theta = \theta_0$ ，那么备择假设就是  $\theta \neq \theta_0$ ；若原假设是  $\theta \geq \theta_0$ ，那么备择假设就是  $\theta < \theta_0$ 。注意原假设一般包含等号。

假设检验的过程就是使用数据作为「证据」试图推翻原假设的过程，这个过程与法官判案的过程类似。在法律中，有所谓的「无罪推定原则」(presumption

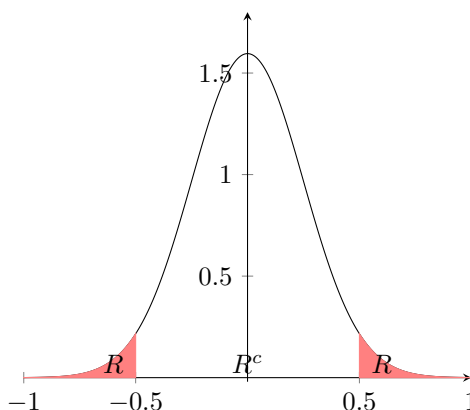


图 1: 拒绝域与第 I 类错误的概率

of innocence)」, 即对于犯罪嫌疑人, 必须先假设其无罪, 原告方有义务提出证据证明其犯罪, 而不得强迫嫌疑人自证其罪。使用以上术语, 即原假设 ( $H_0$ ) 为被告无罪, 备择假设 ( $H_1$ ) 为被告有罪, 假设检验的目的就是使用证据 (样本数据) 试图推翻原假设 (无罪)。如果现有证据可以推翻原假设, 那么我们称为拒绝原假设 (rejecting  $H_0$ ), 即可以认为原假设为假; 而如果现有证据不能推翻原假设, 即没有充足的证据证明原假设为假, 那么我们称不能拒绝原假设 (not rejecting  $H_0$ )。注意「接受原假设 (accepting  $H_0$ )」的说法与「不能拒绝原假设」的说法有细微差别, 如果不能拒绝原假设, 可能是由于我们的证据不够充分, 因而「不能拒绝原假设」的说法更加准确。基于上述原因, 我们一般会把想要推翻的结论放在原假设上。

由于统计方法总会存在误差, 因而基于以上两类假设的推断也会存在犯错的可能性。在假设检验中, 有两种错误可能会发生:

1. 第 I 类错误: 原假设为真, 但是拒绝原假设, 即「弃真错误」;
2. 第 II 类错误: 备择假设为真, 但是接受原假设, 即「取伪错误」。

比如, 如果一个被告本来无罪, 但是错误地判其有罪, 那么就犯了第 I 类错误; 而如果一个被告的确犯罪, 但是却判其无罪, 那么就犯了第 II 类错误。

假设检验一般通过设定一个检验统计量  $T(x)$ , 以及一个拒绝域  $R$ , 当  $T(x) \in R$  时拒绝原假设。如果记  $H_0: \theta \in \Theta_0$ , 而  $H_1: \theta \in \Theta_0^c$ , 那么第 I 类错误, 即原假设为真但是拒绝原假设的概率为:  $P_{\theta \in \Theta_0}(T(x) \in R)$ 。

**例 1.** 如果样本  $x_i \sim N(\mu_0, 1)$  *i.i.d.*,  $i = 1, \dots, N$ ,  $\mu$  为未知总体参数。如果原假设为  $H_0: \mu_0 = 0$ , 备择假设为  $H_1: \mu_0 \neq 0$ , 那么  $\Theta_0 = \{0\}$ 。令检验统计量  $T(x) = \bar{x}$ , 如果在原假设条件下, 即假设  $\mu_0 = 0$ , 那么  $\bar{x} \sim N(0, \frac{1}{N})$ , 即如果原假设成立, 那么样本均值应该分布在 0 附近。而如果我们看到了样本均值距离 0 比较远, 那么说明原假设有可能是不成立的。如图 (1) 所示, 如果取拒绝

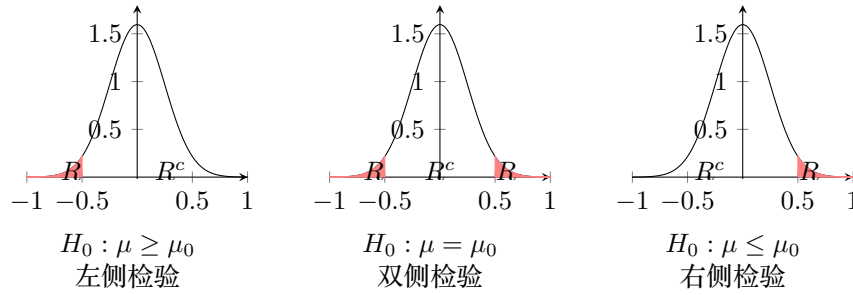


图 2: 单侧检验与双侧检验

域为  $R = (-\infty, -0.5) \cup (0.5, \infty)$ , 那么:

$$\begin{aligned}
 P_{\theta_0=0}(\bar{x} \in R) &= P_{\theta_0=0}(|\bar{x}| > 0.5) \\
 &= P_{\theta_0=0}\left(\left|\frac{\bar{x}}{\sqrt{\frac{1}{N}}}\right| > \frac{0.5}{\sqrt{\frac{1}{N}}}\right) \\
 &= 2\left(1 - \Phi\left(\frac{0.5}{\sqrt{\frac{1}{N}}}\right)\right)
 \end{aligned}$$

如果令  $N = 16$ , 那么  $P_{\theta_0=0}(\bar{x} \in R) = 2(1 - \Phi(2)) \approx 4.56\%$ 。注意以上概率是我们在原假设的假设上进行的计算的, 这意味着, 如果原假设成立, 那么得到均值落在拒绝域, 即  $\bar{x} \in R = (-\infty, -0.5) \cup (0.5, \infty)$  的概率为 4.56%。这样, 如果我们采取「如果均值落入拒绝域就拒绝原假设」这一策略, 那么在原假设的条件下, 我们犯错的概率就是 4.56%, 即犯第 I 类错误的概率为 4.56%。

此外, 如图 (2) 所示, 根据原假设的不同, 检验还可以分为单侧检验和双侧检验。在上例中我们讨论的是双侧检验, 并计算了给定拒绝域的情况下犯第 I 类错误的概率。然而如果原假设是不等于号, 我们通常不能精确计算犯第 I 类错误的概率。此时, 值得关注的是犯第 I 类错误的概率的上界, 即  $\sup_{\theta \in \Theta_0} P_{\theta}(T(x) \in R)$ 。

**例 2.** 如果样本  $x_i \sim N(\mu_0, 1)$  *i.i.d.*,  $i = 1, \dots, N$ ,  $\mu$  为未知总体参数。如果原假设为  $H_0: \mu_0 \leq 0$ , 那么当  $\bar{x}$  足够大时, 可以拒绝原假设。在原假设的条件下,

当  $\mu_0 = \mu < 0$  时, 如果取拒绝域为  $R = (0.5, \infty)$ , 那么:

$$\begin{aligned} P_\mu(\bar{x} \in R) &= P_\mu(\bar{x} > 0.5) \\ &= P_\mu\left(\frac{\bar{x} - \mu}{\sqrt{\frac{1}{N}}} > \frac{0.5 - \mu}{\sqrt{\frac{1}{N}}}\right) \\ &= 1 - \Phi\left(\frac{0.5 - \mu}{\sqrt{\frac{1}{N}}}\right) \end{aligned}$$

注意以上概率随着  $\mu$  的增加而增加, 由于  $\mu \in \Theta_0 = (-\infty, 0]$ , 因而其上界:

$$\sup_{\theta \in \Theta_0} P_\theta(T(x) \in R) = 1 - \Phi\left(\frac{0.5 - 0}{\sqrt{\frac{1}{N}}}\right) = 1 - \Phi\left(\frac{0.5}{\sqrt{\frac{1}{N}}}\right)$$

当  $N = 16$  时, 上式为  $1 - \Phi(2) \approx 2.28\%$ , 即在原假设  $H_0: \mu_0 \leq 0$  的假设下, 错误拒绝原假设的概率上界为  $2.28\%$ , 或者等价的, 犯第 I 类错误的概率上界为  $2.28\%$ 。

在假设检验中, 我们希望控制犯第 I 类错误的概率进行推断, 即给定一个  $\alpha$ , 找到一个拒绝域  $R_\alpha$ , 使得  $\sup_{\theta \in \Theta_0} P_\theta(T(x) \in R_\alpha) \leq \alpha$ 。如此, 我们便保证了使用拒绝域  $R_\alpha$  进行假设检验, 犯第 I 类错误的概率不超过  $\alpha$ 。我们称  $\alpha$  为 **显著性水平 (level of significance)**, 一般取  $\alpha = 0.01, 0.05, 0.1$ , 而  $R_\alpha$  为一个区间, 区间的断点成为临界值 (critical value)。更小的  $\alpha$  代表我们对犯第 I 类错误更加不能容忍, 因而我们会更大的概率接受原假设, 即拒绝域也会越小。

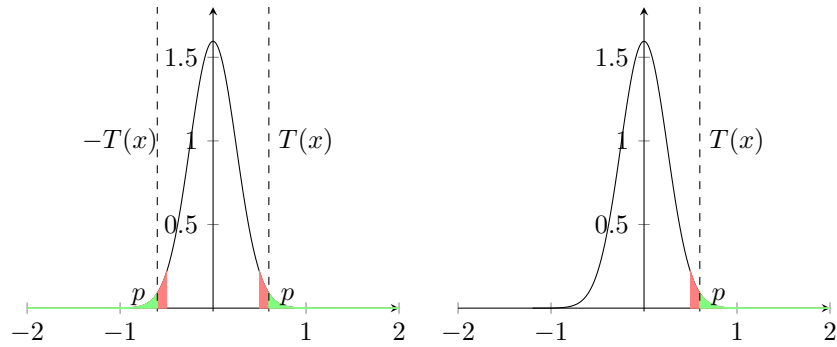
此外, 我们还可以定义  $p$  值的概念。 $p$  值指的是, 给定检验统计量  $T(x) = t$ , 在原假设的条件下, 能够取到  $t$  或者比  $t$  更加极端的值的概率。或者等价的,  $p$  值可以被定义为能够拒绝原假设的最小的显著性水平, 即给定  $T(x) = t$ :

$$p = \inf\{\alpha \in (0, 1) : t \in R_\alpha\}$$

由于检验统计量  $T(x)$  为随机变量, 因而  $p(x) = \inf\{\alpha \in (0, 1) : T(x) \in R_\alpha\}$  也是一个随机变量。当  $p \leq 0.01, 0.05, 0.1$  时, 可以拒绝原假设。如图 (3) 所示, 绿色区域面积即  $p$  值。

综上, 一般的假设检验的步骤可以总结如下:

1. 确定原假设  $H_0$  和备择假设  $H_1$ ;
2. 找到一个检验统计量  $T(x)$  (通常为基准统计量);
3. 确定检验统计量  $T(x)$  在原假设  $H_0$  下的分布;
4. 设定显著性水平  $\alpha$ , 并根据  $\alpha$  确定拒绝域  $R_\alpha$ , 若  $T(x) \in R_\alpha$  则在  $\alpha$  的显著性水平下拒绝原假设; 或者根据  $T(x)$  计算  $p$  值, 若  $p < \alpha$  则拒绝原

图 3:  $p$  值的定义

假设。

**例 3.** 如果样本  $x_i \sim N(\mu, \sigma^2)$   $i.i.d, i = 1, \dots, N$ , 为了检验  $H_0: \mu = \mu_0$ , 在  $H_0$  的假定下,  $\bar{x} \sim N(\mu_0, \frac{\sigma^2}{N})$ , 因而可以构建统计量:

$$t = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} \sim t_{N-1}$$

由于是双侧检验, 因而临界值为  $t_{\alpha/2}$ , 拒绝域为  $R_\alpha = (-\infty, -t_{\alpha/2}) \cup (t_{\alpha/2}, \infty)$ , 当  $|t| > t_{\alpha/2}$  时拒绝原假设, 即认为  $\mu \neq \mu_0$ , 否则不能拒绝原假设。

**例 4.** 如果样本  $x_{1i} \sim N(\mu_1, \sigma_1^2)$   $i.i.d, i = 1, \dots, N_1$ ,  $x_{2i} \sim N(\mu_2, \sigma_2^2)$   $i.i.d, i = 1, \dots, N_2$ , 且两个样本独立。为了检验  $H_0: \sigma_1^2 = \sigma_2^2$ , 在  $H_0$  的假定下,

$$\frac{\frac{(N_1-1)s_1^2}{\sigma_1^2} / (N_1 - 1)}{\frac{(N_2-1)s_2^2}{\sigma_2^2} / (N_2 - 1)} = \frac{s_1^2}{s_2^2} \sim F(N_1 - 1, N_2 - 1)$$

因而其拒绝域为  $R_\alpha = (0, F_{\alpha/2}) \cup (F_{1-\alpha/2}, \infty)$

**例 5.** 根据 2009 年中国城镇住户调查, 在 37480 户家庭中, 已知家庭年收入均值为 54157.63 元, 标准差为 38533.96 元, 请问在 1%、5% 的显著性水平下, 是否可以认为家庭年收入均值为 53000 元? 在这里, 原假设为  $H_0: \mu_0 = 53000$ , 在原假设条件下, 我们有:

$$z = \frac{\bar{x} - \mu_0}{\sqrt{\frac{s^2}{N}}} = \frac{\sqrt{37480}(\bar{x} - 53000)}{38533.96} \stackrel{a}{\sim} N(0, 1)$$

计算可得  $z = 5.82$ , 查表得到在 1% 的显著性水平下, 拒绝域为  $(-\infty, -2.58) \cup (2.58, \infty)$ , 显然  $z$  在拒绝域范围内, 因而可以拒绝原假设。在 1% 的显著性水平下拒绝原假设, 因而在 5% 的显著性水平下必然也拒绝原假设。

**例 6.** 根据 2013 年中国家庭金融调查, 样本 7711 户家庭中, 有 6% 的家庭有信用卡, 请问在 5% 的显著性水平下, 我们是否可以认为我国家庭持有信用卡的比例超过了 5%? 为了解决这一问题, 原假设为:  $H_0: p \leq 5\%$ , 在原假设的条件下, 我们有:

$$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{N}}} = \frac{\hat{p} - 5\%}{\sqrt{\frac{5\%(1-5\%)}{7711}}} \stackrel{a}{\sim} N(0, 1)$$

带入计算可得检验统计量  $z = 4.03$ 。由于是右侧检验, 因而拒绝域为  $(\Phi^{-1}(0.95), \infty)$ , 即  $(1.65, \infty)$ ,  $4.93 > 1.65$ , 因而拒绝原假设, 可以认为我国家庭持有信用卡的比例超过了 5%。

**例 7.** 在 2009 年中国城镇住户调查中, 共有 23440 位 20-50 岁的男性, 以及 21184 位 20-50 岁的女性。已知男性年平均收入为 28367.96 元, 标准差为 21811.88 元; 女性年平均收入为 20145.77 元, 标准差为 16541.08 元。如果假设男女收入独立, 请问在 5% 的显著性水平下, 是否可以认为男女收入差异没有超过 10000 元? 这里, 原假设为  $H_0: \mu_1 - \mu_2 \geq 10000$ , 在原假设的条件下, 我们有:

$$z = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \stackrel{a}{\sim} N(0, 1)$$

由于  $\bar{x}_1 - \bar{x}_2 = 28367.96 - 20145.77 = 8222.19$ ,  $\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2} = \frac{21811.88^2}{23440} + \frac{16541.08^2}{21184} = 33212.60$ ,  $\mu_1 - \mu_2 = 10000$ , 因而  $z = -9.76$ 。由于是左侧检验, 因而拒绝域为  $(-\infty, -1.65)$ , 可以拒绝原假设, 即可以认为男女收入差异没有超过 10000 元。

## 2 评价检验的方法

根据以上的假设检验步骤, 我们知道假设检验可以控制犯第 I 类错误的概率, 然而对于假设检验而言, 还有犯第 II 类错误的可能性, 即当我们无法拒绝原假设的时候, 备择假设可能是成立的。为了研究第 II 类错误的概率, 我们引入假设检验的势 (power) 的概念。

**定义 1.** 对于一个假设检验及其拒绝域  $R$ , 检验的**势函数** (power function) 即给定  $\theta$  拒绝原假设的概率, 即  $\beta(\theta) = P_\theta(T(x) \in R)$ 。

注意在以上定义中我们并没有限定  $\theta \in \Theta_0$  或者  $\theta \in \Theta_0^c$ , 因而理想的情况下, 当  $\theta \in \Theta_0$  时,  $\beta(\theta)$  应该接近于 0, 而当  $\theta \in \Theta_0^c$  时,  $\beta(\theta)$  应该接近于 1。当  $\theta \in \Theta_0^c$ , 即备择假设为真时, 没有拒绝原假设的概率, 即  $1 - \beta(\theta)$ , 即犯第 II 类错误的概率。

**例 8.** 若样本  $x_i \sim N(\mu, \sigma^2)$  *i.i.d.*,  $i = 1, \dots, N$ , 且  $\sigma^2$  已知, 设原假设为  $H_0: \mu \leq \mu_0$ , 备择假设为  $H_1: \mu > \mu_0$ , 那么可以构建检验统计量为:

$$T(x) = \frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} \sim N(0, 1)$$

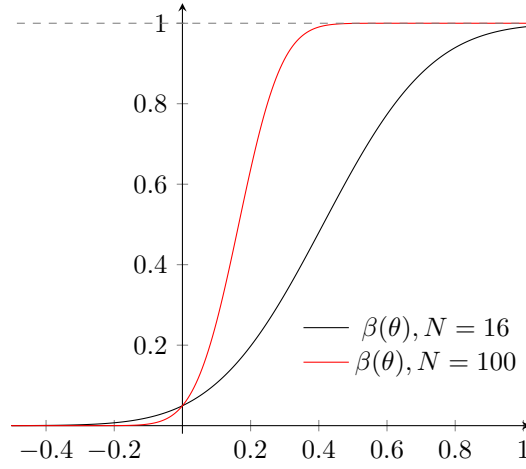


图 4: 势函数

在原假设的条件下, 若给定  $\alpha = 0.05$ , 那么  $R_{0.05} = (z_{0.95}, \infty)$ 。因而, 给定  $\mu$ , 势函数为:

$$\begin{aligned}
 \beta(\theta) &= P_{\mu}(T(x) \in R_{0.05}) \\
 &= P_{\mu}\left(\frac{\bar{x} - \mu_0}{\sqrt{\frac{\sigma^2}{N}}} > z_{0.95}\right) \\
 &= P_{\mu}\left(\frac{\bar{x} - \mu}{\sqrt{\frac{\sigma^2}{N}}} > z_{0.95} + \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{N}}}\right) \\
 &= 1 - \Phi\left(z_{0.95} + \frac{\mu_0 - \mu}{\sqrt{\frac{\sigma^2}{N}}}\right)
 \end{aligned}$$

图 (4) 给出了当  $\mu_0 = 0$ ,  $N = 16$ ,  $\sigma^2 = 1$  时的势函数。注意当  $\mu \rightarrow -\infty$  时,  $\beta(\mu) \rightarrow 0$ ; 当  $\mu = \mu_0$  时,  $\beta(\mu) = 0.05$ ; 而当  $\mu \rightarrow \infty$  时,  $\beta(\mu) \rightarrow 1$ , 且  $\beta(\theta)$  是  $\mu$  的单调递增函数。这意味着, 在当原假设为真时, 即  $\mu \leq \mu_0$  时, 拒绝原假设的概率总是小于等于  $\alpha = 0.05$  的, 这与我们假设检验控制第 I 类错误的概率是一致的。而当  $\mu > \mu_0$  时, 随着  $\mu$  的增大, 犯第 II 类错误的概率也随之降低。此外, 注意到, 犯第 II 类错误的概率是随着样本量  $N$  的增大而减小的。

观察图 (4) 我们会发现, 不管样本量  $N$  再大, 当真值  $\theta > \theta_0$ , 但是差异很小时, 第 II 类错误概率  $1 - \beta(\theta)$  仍然会无线趋向于  $1 - \alpha$ 。因而当备择假设为真, 但是真值与原假设非常接近时, 仍然需要样本量非常大才能够正确的拒绝原假设。

介于此种情况, 我们可以引入无差异区域 (indifference region), 即虽然备择假设为真, 但是  $\theta$  与  $\theta_0$  的差异足够的小, 我们认为在这个区域里面错误接

受原假设也是可以接受的。例如，如果我们设计实验研究 wifi 会不会致癌，即看照射组合非照射组的实验对象患癌症的概率是否显著大于 0，即  $H_0: \mu \leq 0$ 。假设 wifi 的确会致癌，但是致癌的概率充分的接近于 0，根据图 (4) 我们会发现，如果我们想要正确的拒绝原假设，需要非常大的样本量才能保证以一个比较大的概率  $\beta$  拒绝原假设。然而如果我们认为，致癌概率在一定的范围内，比如  $[0, \Delta)$  内，是可以接受的，那么我们可以据此设计样本量，保证以一个确定的概率拒绝原假设。

如果我们希望，当  $\mu = \Delta$  时，至少以  $\beta$  的概率拒绝原假设，那么

$$\beta = 1 - \Phi \left( z_\alpha + \frac{0 - \Delta}{\sqrt{\frac{\sigma^2}{N}}} \right)$$

从而

$$N = \left[ \frac{\sigma}{\Delta} (z_\alpha - z_{1-\beta}) \right]^2$$

**例 9.** 在上例中，记实验组 (wifi 照射) 得癌症的概率为  $p_1$ ，对照组 (无 wifi 照射) 得癌症的概率为  $p_2$ ，我们关注的关键变量为  $\mu = p_1 - p_2$ ，原假设为  $H_0: \mu \leq 0$ 。如果拒绝原假设，即可以认为 wifi 照射会致癌。如果假设两组的样本量相同，那么：

$$\text{Var}(\hat{\mu}) = \text{Var}(\hat{p}_1 - \hat{p}_2) = \text{Var}(\hat{p}_1) + \text{Var}(\hat{p}_2) = \frac{p_1(1-p_1) + p_2(1-p_2)}{N}$$

在原假设下，可以认为  $p_1 = p_2$ ，那么  $\text{Var}(\hat{\mu}) = \frac{2p_1(1-p_1)}{N}$ 。如果取  $\Delta = 0.001$ ， $\beta = 0.8$ ，记我们希望当 wifi 致癌的概率为千分之一时，我们希望以 80% 的概率拒绝原假设，取  $\alpha = 0.05$ ，那么所需样本量为：

$$N = \left[ \frac{2\sqrt{p_1(1-p_1)}}{\Delta} (1.65 - 0.15) \right]^2 = 4 \left( \frac{1.65 - 0.15}{0.0001} \right)^2 [p_1(1-p_1)]$$

如果自然条件下，癌症发病率为 0.2%，即当  $p_1 = 0.002$  时，上述样本量大概需要  $N = 17964$  个。而反过来，如果样本量足够大，那么  $\mu$  与  $\mu_0$  的一些非常细微的差别也足以导致统计上的显著性，尽管很多时候这种细微的差别几乎没有现实意义。因而，即使数据中可以得到统计上显著的结果，特别是在样本量非常大的情况下，我们仍然要注意这些结果是不是有「经济显著性 (economic significancy)」，即这些差别在现实中是不是足以引起重视。

### 3 构造假设检验的方法

在以上两节中我们介绍了假设检验的一般概念和思路。我们知道，如果在原假设  $H_0$  的条件下得到检验统计量及其抽样分布，我们就可以使用上节中给出的步骤进行假设检验。尽管上一节中我们讨论了单个样本、多个样本均值的



假设检验，然而很多时候我们可能希望对不止一个参数进行假设检验，或者对参数的函数进行假设检验。一般的，记我们的原假设为：

$$H_0 : C(\theta) = 0$$

其中  $\theta \in \mathbb{R}^k$ ,  $C(\theta) \in \mathbb{R}^r$ , 且  $C(\theta)$  为  $\theta$  的连续可微函数。那么一般的，我们可以通过如下两个方法构造假设检验。

### 3.1 Wald 检验

如果对于  $\theta$  的估计，我们已经有  $\hat{\theta} \sim N(\theta, \Sigma)$ ，那么使用 Delta 方法：

$$C(\hat{\theta}) = C(\theta) + \ddot{C}(\hat{\theta} - \theta) + o(|\hat{\theta} - \theta|)$$

其中  $\ddot{C} = \partial C(\theta) / \partial \theta$  为  $r \times k$  的矩阵，且假设  $\text{rank}(\ddot{C}) = r$ ，那么在原假设的条件下，：

$$C(\hat{\theta}) = \ddot{C}(\hat{\theta} - \theta) + o_p(1) \sim N(0, \ddot{C}\Sigma\ddot{C}')$$

进而我们可以构建检验统计量：

$$C'(\hat{\theta}) [\ddot{C}\Sigma\ddot{C}']^{-1} C(\hat{\theta}) \sim \chi_r^2$$

因而可以使用以上检验统计量对原假设进行假设检验。

### 3.2 似然比检验

在原假设的条件下，如果可以使用极大似然估计对  $\theta$  进行估计，那么我们可以 在  $H_0$  的约束下对  $\theta$  进行估计，即：

$$\begin{aligned} \tilde{\theta} &= \arg \max_{\theta} L(\theta|x) \\ s.t. & C(\theta) = 0 \end{aligned}$$

与此同时，我们还可以估计无约束的极大似然估计：

$$\hat{\theta} = \arg \max_{\theta} L(\theta|x)$$

由于  $\tilde{\theta}$  是在约束条件下得到的，而  $\hat{\theta}$  是在无约束的条件下得到的，因而  $L(\hat{\theta}|x) \geq L(\tilde{\theta}|x)$ 。如果原假设成立，即  $C(\theta) = 0$ ，那么  $\tilde{\theta}$  是和  $\hat{\theta}$  应该充分接近，因而  $L(\hat{\theta}|x)$  和  $L(\tilde{\theta}|x)$  也应该充分接近；但是如果  $L(\hat{\theta}|x) > L(\tilde{\theta}|x)$  成立，那么

我们可以认为  $C(\theta) = 0$  是不成立的。实际上, 我们可以得到:

$$LR = 2 \left[ L(\hat{\theta}|x) - L(\tilde{\theta}|x) \right] \sim \chi_r^2$$

因而我们可以根据以上结论对  $C(\theta) = 0$  进行检验。

**例 10.** 如果  $x_i \sim P(\lambda)$ ,  $i = 1, \dots, N$ , 如果原假设为  $H_0: \lambda = 1$ , 那么无约束时,  $\hat{\lambda} = \bar{x}$ , 而有约束时,  $\tilde{\lambda} = 1$ 。其似然函数分别为:

$$\begin{aligned} L(\hat{\lambda}|x) &= \sum_{i=1}^N \left[ x_i \ln(\hat{\lambda}) - \ln(x_i!) - \hat{\lambda} \right] = \sum_{i=1}^N [x_i \ln(\bar{x}) - \ln(x_i!) - \bar{x}] \\ L(\tilde{\lambda}|x) &= \sum_{i=1}^N \left[ x_i \ln(\tilde{\lambda}) - \ln(x_i!) - \tilde{\lambda} \right] = \sum_{i=1}^N [x_i \ln(1) - \ln(x_i!) - 1] \end{aligned}$$

因而检验统计量为:

$$\begin{aligned} LR &= 2 \left[ \sum_{i=1}^N [x_i \ln(\bar{x}) - \ln(x_i!) - \bar{x}] - \sum_{i=1}^N [x_i \ln(1) - \ln(x_i!) - 1] \right] \\ &= 2 \left[ \sum_{i=1}^N [x_i \ln(\bar{x}) - \bar{x} + 1] \right] \\ &= 2N [\ln(\bar{x}) \bar{x} - \bar{x} + 1] \sim \chi_1^2 \end{aligned}$$

## 习题

- (程序题) 生成一系列正态分布  $x_i \sim N(\mu, \sigma^2)$ , 给定不同的  $\mu \in [-2, 2]$ , 在  $H_0: \mu = 0$  的条件下, 对于固定的  $\mu$ , 重复 1000 次假设检验的过程, 并计算对于每一个  $\mu$ , 在  $H_0: \mu = 0$  的条件下拒绝原假设的比率 (power function)。
- (程序题) 对于比例的检验, 在  $H_0: p = p_0$  的原假设下, 我们有如下两个渐进等价的方法:

$$\frac{\hat{p} - p}{\sqrt{\frac{p_0(1-p_0)}{N}}} \stackrel{a}{\sim} N(0, 1)$$

以及:

$$\frac{\hat{p} - p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{N}}} \stackrel{a}{\sim} N(0, 1)$$

分别模拟两种检验方法在小样本 ( $N = 15$ ) 和大样本 ( $N = 50$ ) 条件下的势函数, 并进行比较。

- (程序题) 请使用对数似然比检验, 在  $H_0: p = p_0$  的原假设下, 写出对样本比例的检验统计量, 并使用数值模拟计算出其势函数, 并于上题中的势

函数进行比较。

## 参考文献

- [1] Casella, G., Berger, R.L., 2002. Statistical inference. Duxbury Pacific Grove, CA.
- [2] Shao, J., 2007. Mathematical Statistics, 2nd ed. Springer, New York.