基础数理统计

第十章 假设检验和 p 值

假设检验引例-Human sex ratio

最早的假设检验应用要追溯到 1700s，John Arbuthnot 研究了人类男性和女性出生比例是否相等。

- Arbuthnot 检查了伦敦从 1629 年到 1710 年连续 82 年的出生记录；
- 发现这 82 年的每一年，伦敦的男性出生数都高于女性；
- 如果男女出生比率应该是相等的，那么这一结果出现的可能性只有 $0.5^{82} \approx 1/4, 8360, 0000, 0000, 0000, 0000, 0000.$
- 这一可能性是如此之小，Arbuthnot 认为产生这一结果不是由于随机性，而是客观证据表明: "From whence it follows, that it is Art, not Chance, that governs."
- 统计上来说，基于伦敦从 1629 年到 1710 年连续 82 年的出生记录来看，我们不认为男性女性出生比例相同。

假设检验引例-女士品茶

女士品茶是统计学历史上最著名的试验，充分展现了统计智慧:

- 20 世纪 20 年代后期，在英国剑桥一个夏日的午后，一群大学的绅士和他们的夫人们享用着下午茶。在品茶过程中，一位女士 (Dr. Muriel Bristol) 坚称：**把茶加进奶里，或把奶加进茶里，不同的做法，会使茶的味道品起来不同。**
- 在座的一个身材矮小、戴着厚眼镜、下巴上蓄着的短尖髯开始变灰的先生 (Ronald Fisher)，却不这么看，他对这个问题很感兴趣。他兴奋地说道：**“让我们来检验这个命题吧！”**。

如何检验女士品茶?

思考

假设做十次独立的试验，女士猜对多少杯你会认为她真的有鉴别能力?

- 即使女士真的有鉴别能力，她可能因随机性没有猜对很多杯;
- 即使女士没有鉴别能力，她也可能因为运气因素猜对很多杯;
- 不同的学科对这种鉴别能力的要求也有所不同。

参数假设检验

给定一组样本 $X_1, \cdots, X_n \sim f(x, \theta)$，参数假设检验问题是根据已有数据推断

$$H_0: \ \theta \in \Theta_0 \ \ v.s. \ \ H_1: \ \theta \in \Theta_1.$$

其中，

- H_0 称为原假设 (the null hypothesis)
- H_1 称为备择假设 (the alternative hypothesis)

检验拒绝域

样本 $X_1, \cdots, X_n$ 的所有可能结果称为**样本空间** Ω, 假设检验是定义 Ω 的一个子集 $\mathcal{A}$,

$$\text{如果观察样本} \quad (X_1, \cdots, X_n) \in \mathcal{A},$$

那么我们就拒绝原假设, $\mathcal{A}$ 称为拒绝域。

备注:

假设检验本质上就是构造合适的拒绝域 $\mathcal{A}$。

检验统计量

通常，拒绝域会通过检验统计量来表示，

$$\mathcal{A} = \{(x_1, \cdots, x_n) \in \Omega : T(x_1, \cdots, x_n) > c\},$$

其中

- $T = T(X_1, \cdots, X_n)$ 就是检验统计量。
- c 称为临界值。

假设检验即是构造合适的检验统计量和找到合适的临界值。

假设检验的两类错误

	H_0 为真	H_1 为真
$(X_1,\cdots,X_n)\in\mathcal{A}^c$	正确决策	第二类错误
$(X_1,\cdots,X_n)\in\mathcal{A}$	第一类错误	正确决策

- 第一类错误 (假阳性)：H_0 被错误地拒绝

$$\text{第一类错误} = P((X_1,\cdots,X_n)\in\mathcal{A}|H_0).$$

- 第二类错误 (假阴性)：H_0 被错误地接受

$$\text{第二类错误} = P((X_1,\cdots,X_n)\notin\mathcal{A}|H_1).$$

两类错误的平衡

- 为了控制第一类错误，拒绝域 $\mathcal{A}$ 应该越小越好；例如极端的 $\mathcal{A} = \emptyset$ 可以保证第一类错误发生概率为 0。
- 而为了控制第二类错误，$\mathcal{A}$ 越大越好；例如极端的 $\mathcal{A} = \Omega$ 可以保证第二类错误发生概率为 0。

统计中采用的是控制第一类错误不超过一定的水平，例如 $\alpha = 5\%$。

势函数

定义 1 (势函数)

对于给定的拒绝域 $\mathcal{A}$，势函数定义为

$$\beta(\theta) = P(X \in \mathcal{A} | X_1, \cdots, X_n \sim f(x, \theta)) = P_\theta(X \in \mathcal{A}).$$

定义假设检验的容度为

$$\beta = \sup_{\theta \in \Theta_0} \beta(\theta).$$

如果检验的容度小于等于 α，称检验的水平为 α。

说明:

(1). 检验水平：取决于检验准则；

(2). 显著性水平：通常是每次检验之前确定，允许的第一类错误的概率；

(3). 检验水平 vs 显著性水平：通常检验准则中的临界值是根据显著性水平来确定的，这种情况下，检验水平就是显著性水平；

(4). 检验的容度 vs 检验水平：如果检验容度小于等于 α，称检验水平为 α。

假设

- $\theta = \theta_0$：简单假设
- $\theta > \theta_0$, $\theta < \theta_0$：复合假设
- $H_0 : \theta = \theta_0$ vs $H_1 : \theta \neq \theta_0$：双边检验
- $H_0 : \theta \leq \theta_0$ vs $H_1 : \theta > \theta_0$ 或
  $H_0 : \theta \geq \theta_0$ vs $H_1 : \theta < \theta_0$：单边检验

例子

例 1

假定样本 $X_1, \cdots, X_n \sim N(\mu, 1)$，检验问题为：

$$H_0: \ \mu = 0 \ v.s. \ H_1: \mu \neq 0.$$

假设检验的拒绝域为：

$$\mathcal{A} = \{|\bar{X}| > 1.96/\sqrt{n}\}.$$

势函数为：

$$\begin{aligned}
\beta(\mu) &= P(|\sqrt{n}\bar{X}| > 1.96 | X_1, \cdots, X_n \sim N(\mu, 1)) \\
&= P(\sqrt{n}(\bar{X} - \mu) > 1.96 - \sqrt{n}\mu) \\
&\quad + P(\sqrt{n}(\bar{X} - \mu) < -1.96 - \sqrt{n}\mu) \\
&= 1 - \Phi(1.96 - \sqrt{n}\mu) + \Phi(-1.96 - \sqrt{n}\mu).
\end{aligned}$$

例子–续

例 2 (单边检验)

假定样本 $X_1, \cdots, X_n \sim N(\mu, 1)$，检验问题为：

$$H_0: \ \mu \le 0 \ v.s. \ H_1: \mu > 0.$$

假设检验的拒绝域为：$\mathcal{A} = \{\bar{X} > c\}$。势函数为：

$$\begin{aligned} \beta(\mu) &= P(\bar{X} > c | X_1, \cdots, X_n \sim N(\mu, 1)) \\ &= P(\sqrt{n}(\bar{X} - \mu) > \sqrt{n}(c - \mu)) = 1 - \Phi(\sqrt{n}(c - \mu)) \end{aligned}$$

检验水平

$$\alpha = \sup_{\mu \le 0} \beta(\mu) = \beta(0) = 1 - \Phi(\sqrt{n} c).$$

如果我们设定显著性水平 $\alpha (= 0.05)$, 那么

$$c = \Phi^{-1}(1 - \alpha) / \sqrt{n}.$$

假设检验流程

假设检验问题 $H_0: \theta \in \Theta_0$ v.s. $H_1: \theta \in \Theta_1$ 的检验机制

- 构造检验统计量 T (一般为参数 θ 的点估计);
- 计算 T 的精确或者渐近分布;
- 根据检验问题以及显著性水平确定出拒绝域 $\mathcal{A}$;
- 代入样本判断统计量是否落入拒绝域给出检验结果。(并解释科学意义)

常见检验

假设我们构造出了一元参数 θ 的估计统计量 T,

- 双边检验

$$H_0: \ \theta = \theta_0 \ \text{v.s.} \ H_1: \ \theta \neq \theta_0,$$

拒绝域为 $\mathcal{A} = \{|T - \theta_0| > c_\alpha^{(1)}\}$.

- 单边检验

$$H_0: \ \theta > \theta_0 \ \text{v.s.} \ H_1: \ \theta \leq \theta_0,$$

拒绝域为 $\mathcal{A} = \{T - \theta_0 < c_\alpha^{(2)}\}$.

$$H_0: \ \theta \leq \theta_0 \ \text{v.s.} \ H_1: \ \theta > \theta_0,$$

拒绝域为 $\mathcal{A} = \{T - \theta_0 > c_\alpha^{(3)}\}$.

10.1 Wald 检验

10.1 Wald 检验
10.2 p 值
10.3 χ^2 分布
10.4 多项分布数据的 Pearson χ^2 检验
10.5 置换检验
10.6 似然比检验
10.7 多重假设检验
10.8 拟合优度检验

Wald 检验

令 θ 为尺度参数，$\widehat{\theta}$ 为 θ 的估计，$\widehat{\mathsf{se}}$ 为 $\widehat{\theta}$ 的标准差的估计。

定义 2 (Wald 检验)

考虑检验

$$H_0: \ \theta = \theta_0 \ \ v.s. \ \ H_1: \ \theta \neq \theta_0,$$

假设 $\widehat{\theta}$ 是渐近正态的，

$$\frac{\widehat{\theta} - \theta_0}{\widehat{\mathsf{se}}} \rightsquigarrow N(0, 1),$$

显著水平为 α 的 *Wald* 检验：当 $|W| > z_{\alpha/2}$ 时拒绝 H_0，其中

$$W = \frac{\widehat{\theta} - \theta_0}{\widehat{\mathsf{se}}}.$$

Wald 检验–续

定理 1

渐近的，*Wald* 检验的显著水平为 α，即当 $n \to \infty$ 的时候，

$$P_{\theta_0}(|W| > z_{\alpha/2}) \to \alpha.$$

注：Wald 检验另一个检验统计量为 $W = (\widehat{\theta} - \theta_0)/\mathrm{se}_0$，其中 se_0 是在 $\theta = \theta_0$ 时计算出来的标准差。

Wald 检验的函数

定理 2

假设 θ 的真实值为 $\theta_\star \neq \theta_0$，势函数为：

$$\beta(\theta_\star) \approx 1 - \Phi(z_{\alpha/2} + \frac{\theta_0 - \theta_\star}{\widehat{se}}) + \Phi(-z_{\alpha/2} + \frac{\theta_0 - \theta_\star}{\widehat{se}}).$$

Wald 检验–例子

例 3 (比较两个均值)

给定两组观察样本

$$X_1, \cdots, X_m; \quad Y_1, \cdots, Y_n,$$

检验两组数据的总体均值是否相等？假设检验问题：

$$H_0: \ \mu_1 = \mu_2 \ \text{v.s.} \ H_1: \ \mu_1 \neq \mu_2.$$

等价的：

$$H_0: \ \mu_1 - \mu_2 = 0 \ \text{v.s.} \ H_1: \ \mu_1 - \mu_2 \neq 0.$$

例子–续

构造 $\theta = \mu_1 - \mu_2$ 的估计为：$\widehat{\theta} = \bar{X} - \bar{Y}$.
计算方差：

$$Var(\widehat{\theta}) = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n},$$

其中 σ_1, σ_2 分别是两组数据各自的总体标准差，可令

$$\widehat{se} = \sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}},$$

其中

$$S_1^2 = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \bar{X})^2,\ S_2^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2.$$

检验拒绝域为

$$\left|\frac{\bar{X} - \bar{Y}}{\widehat{se}}\right| > z_{\alpha/2}.$$

配对检验

例 4

给定观察样本

$$(X_1, Y_1), \cdots, (X_n, Y_n).$$

检验两个总体的均值是否相等？假设检验问题：

$$H_0: \ \mu_1 = \mu_2 \ \ v.s. \ \ H_1: \ \mu_1 \neq \mu_2.$$

配对检验–续

构造 $\theta = \mu_1 - \mu_2$ 的估计为：$\widehat{\theta} = \bar{X} - \bar{Y}$.
计算方差：

$$Var(\widehat{\theta}) = \frac{\sigma^2}{n},$$

其中 σ 分别是 $X_i - Y_i$ 的总体标准差，实际中可以用样本的标准差来估计，即

$$\widehat{se} = \sqrt{\frac{S^2}{n}}, \ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - Y_i - (\bar{X} - \bar{Y}))^2.$$

检验拒绝域为

$$\left|\frac{\bar{X} - \bar{Y}}{\widehat{se}}\right| > z_{\alpha/2}.$$

Wald 检验–例子

例 5 (比较两个中位数)

检验两个分布的中位数 ν_1, ν_2 是否相同，$\delta = \nu_1 - \nu_2$ 的非参嵌入式估计为 $\widehat{\delta} = \widehat{\nu}_1 - \widehat{\nu}_2$，这里 $\widehat{\nu}_1$, $\widehat{\nu}_2$ 为样本中位数，$\widehat{\delta}$ 的标准差 $\widehat{se}$ 可以通过 *Bootstrap* 方法得到。*Wald* 检验统计量为 $W = \widehat{\delta}/\widehat{se}$。

Wald 检验–续

定理 3

显著水平为 α 的 *Wald* 检验拒绝 $H_0:\theta=\theta_0$，其对立假设为 $H_1:\theta\neq\theta_0$ 当且仅当 $\theta_0\notin C$，其中

$$C=\left(\widehat{\theta}-z_{\alpha/2}\widehat{se},\widehat{\theta}+z_{\alpha/2}\widehat{se}\right).$$

作业：8(选做)，9(必做)

10.2 p 值

p 值

定义 3 (p 值)

假设对于任意 $\alpha \in (0,1)$，存在检验水平为 α 的检验，它的拒绝域为 $\mathcal{A}_\alpha$，则

$$p\text{ 值} = \inf\{\alpha : T(X_1, X_2, \ldots, X_n) \in \mathcal{A}_\alpha\}.$$

即：p 值是可以拒绝 H_0 的最小检验水平。

p 值–续

非正式地，p 值是拒绝 H_0 的证据强弱的度量；p 值越小，拒绝 H_0 的证据越强。

p 值	证据
< 0.01	很强的拒绝 H_0 的证据
$0.01 \sim 0.05$	较强的拒绝 H_0 的证据
$0.05 \sim 0.10$	较弱的拒绝 H_0 的证据
> 0.1	没有证据可以拒绝 H_0

说明：p 值很大不是保留 H_0 的强证据。p 值很大有两个原因 (1)H_0 为真；(2)H_0 假但势函数很低。

p 值计算

定理 4

假设显著性水平为 α 的检验的形式为

$$\text{拒绝 } H_0 \text{ 当且仅当 } T(X_1, X_2, \ldots, X_n) \geq c_\alpha,$$

则

$$p\text{ 值} = \sup_{\theta \in \Theta_0} P_\theta(T(X_1, X_2, \ldots, X_n) \geq T(x_1, x_2, \ldots, x_n)),$$

其中，$(x_1, x_2, \ldots, x_n)$ 是 $(X_1, X_2, \ldots, X_n)$ 的观测值。如果 $\Theta_0 = \{\theta_0\}$，则

$$p\text{ 值} = P_{\theta_0}(T(X_1, X_2, \ldots, X_n) \geq T(x_1, x_2, \ldots, x_n)).$$

p 值计算

定理 5

令 $\omega = (\widehat{\theta} - \theta_0)/\widehat{se}$ 表示 *Wald* 统计量的观测值。p 值由下式给出：

$$p\text{ 值} = P_{\Theta_0}(|W| > \omega) \approx P(|Z| > \omega) = 2\Phi(-|\omega|),$$

其中 $Z \sim N(0,1)$。

p 值分布

定理 6

如果检验统计量服从连续分布，则在 $H_0 : \theta = \theta_0$ 下，p 值服从均匀分布 $U(0,1)$。因此，如果当 p 值小于 α 时拒绝 H_0，那么犯第一类错误的概率为 α。

例 6

比较两个均值的例子中，如果

$$W = |\frac{\bar{X} - \bar{Y}}{\widehat{se}}| = 3.78,$$

则 p 值为 $p = P(|Z| > 3.78)$，这里 $Z \sim N(0,1)$，故 p 值为 0.0002.

作业：5(必做)，6(选做)

10.3 χ^2 分布

定义 4 (χ^2 分布)

对于独立同分布的标准正态分布

$$X_1, \cdots, X_n \ i.i.d \ \sim N(0,1),$$

称统计量

$$T_n = X_1^2 + \cdots + X_n^2,$$

满足的分布为 χ^2 分布，其中 n 为自由度, 记为 $T_n \sim \chi_n^2$.

χ^2 分布的性质

自由度为 n 的 χ^2 分布的密度函数为：

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{n/2-1} e^{-x/2}, \ x \geq 0,$$

其中 $\Gamma(\cdot)$ 为数学中的 Γ 函数。

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt \ (x > 0); \ \Gamma(x+1) = x\Gamma(x).$$

对于标准正态分布 $X \sim N(0,1)$，

$$\mathrm{E}(X^2) = 1, \ \mathrm{E}(X^4) = 3,$$

由此，我们可以得到自由度为 n 的卡方分布的期望为 n，方差为 $2n$，即

$$\mathrm{E}(T_n) = n, \ \mathrm{V}(T_n) = 2n.$$

10.4 多项分布数据的 Pearson χ^2 检验

问题

$X = (X_1, X_2, \ldots, X_k)$ 服从多项分布 Multinomial$(n, \boldsymbol{p})$，令 $\boldsymbol{p}_0 = (p_{01}, p_{02}, \ldots, p_{0k})^\top$ 是某一固定的向量，假设希望检验

$$H_0 : \boldsymbol{p} = \boldsymbol{p}_0 \ \text{ vs } \ H_1 : \boldsymbol{p} \neq \boldsymbol{p}_0.$$

Pearson χ^2 统计量

定义 5 (Pearson χ^2 统计量)

Pearson χ^2 统计量为

$$T = \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} = \sum_{i=1}^{k} \frac{(X_j - E_j)^2}{E_j}$$

其中 $E_j = E(X_j) = np_{0j}$ 是 X_j 在 H_0 下的期望值。

定理 7

在 H_0 下，$T \rightsquigarrow \chi^2_{k-1}$。$\chi^2$ 检验的检验水平为 α 的拒绝域为：

$$T > \chi^2_{k-1}(\alpha).$$

p 值为 $P(\chi^2_{k-1} > t)$，其中 t 是检验统计量的观测值。

渐近分布的证明

性质 1

设 $X \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$, A 为对称矩阵，且 $rank(A) = r$. 则二次型 $\dfrac{X^\top AX}{\sigma^2} \sim \chi^2(r)$ 当且仅当 $A^2 = A$.(A 为对称幂等矩阵)。

A 为对称矩阵，且 $\mathrm{rank}(A) = r$. 则存在正交阵 Γ, 使得

$$\Gamma^\top A\Gamma = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_r, 0, \cdots, 0),$$

其中 $\lambda_1, \ldots, \lambda_r$ 为 A 的非零特征根. 令 $Y = \Gamma^\top X$, 则 $Y \sim N_n(\mathbf{0}, \sigma^2 \boldsymbol{I}_n)$ 且

$$\xi = \frac{X^\top AX}{\sigma^2} = \frac{\sum_{i=1}^{r} \lambda_i Y_i^2}{\sigma^2}.$$

渐近分布的证明–续

证明.

a. “充分性”： A 为对称幂等矩阵, 则 $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$，故 $\xi \sim \chi^2(r)$;

b. “必要性”: $\chi^2(k)$ 的特征函数为 $(1-2it)^{-k/2}$. 一方面，ξ 的特征函数为 $\prod_{j=1}^r (1-2it\lambda_j)^{-1/2}$；另一方面，$\xi \sim \chi^2(r)$, 特征函数为 $(1-2it)^{-r/2}$. 由等式 $\prod_{j=1}^r (1-2it\lambda_j)^{-1/2} = (1-2it)^{-r/2}$ 可得到 $\lambda_1 = \lambda_2 = \cdots = \lambda_r = 1$，故 A 为对称幂等矩阵。

$\square$

渐近分布的证明–续

令 $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n$ 为 i.i.d 服从 Multinomial$(1, \boldsymbol{p}_0)$，则根据中心极限定理，我们有

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \boldsymbol{Y}_i - n\boldsymbol{p}_0\right) \rightsquigarrow N_k(\mathbf{0}, \Sigma),$$

这里 $\Sigma = \mathrm{diag}\{\boldsymbol{p}_0\} - \boldsymbol{p}_0\boldsymbol{p}_0^\top$. 注意到

$$\frac{1}{\sqrt{n}}\begin{pmatrix} X_1 - np_{01} \\ X_2 - np_{02} \\ \cdots \\ \cdots \\ X_k - np_{0k} \end{pmatrix} \overset{d}{=} \frac{1}{\sqrt{n}}\left(\sum_{i=1}^n \boldsymbol{Y}_i - n\boldsymbol{p}_0\right).$$

渐近分布的证明–续

$$
\begin{aligned}
T &= \sum_{j=1}^{k} \frac{(X_j - np_{0j})^2}{np_{0j}} \\
&= \left(\frac{1}{\sqrt{n}} \begin{pmatrix} X_1 - np_{01} \\ X_2 - np_{02} \\ \cdots \\ \cdots \\ X_k - np_{0k} \end{pmatrix} \right)^{\top} \mathrm{diag}(\frac{1}{\boldsymbol{p}_0}) \left(\frac{1}{\sqrt{n}} \begin{pmatrix} X_1 - np_{01} \\ X_2 - np_{02} \\ \cdots \\ \cdots \\ X_k - np_{0k} \end{pmatrix} \right) \\
&\rightsquigarrow \boldsymbol{Z}^{\top} \left(\Sigma^{1/2} \mathrm{diag}(\frac{1}{\boldsymbol{p}_0}) \Sigma^{1/2} \right) \boldsymbol{Z}.
\end{aligned}
$$

$\boldsymbol{Z} \sim N_k(\mathbf{0}, \boldsymbol{I}_k)$。

渐近分布的证明–续

注意到

$$A = \Sigma^{1/2}\mathrm{diag}(\frac{1}{\boldsymbol{p}_0})\Sigma^{1/2} = \boldsymbol{I}_k - \mathrm{diag}(\sqrt{\boldsymbol{p}_0})\mathrm{diag}(\sqrt{\boldsymbol{p}_0})^\top,$$

则 A 是秩为 $k-1$ 的对称幂等矩阵。

- χ^2 检验可以被直接用来检验一组数据是否满足某个离散分布；
- 如果是连续分布，我们可以离散化然后再使用卡方检验，是拟合优度检验。

例子

例 7 (Mendel 的豌豆)

Mendel 把饱满的黄颜色豌豆和皮皱的绿颜色豌豆杂交，后代有四种可能：饱满黄颜色，皮皱黄颜色，皮皱绿颜色，饱满绿颜色。每一种类型的个数服从概率为 $\boldsymbol{p} = (p_1, p_2, p_3, p_4)$ 的多项分布。遗传理论预测

$$\boldsymbol{p}_0 = (\frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16}).$$

在 $n = 556$ 次试验中，观察到

$$X = (315, 101, 108, 32).$$

将检验 $H_0 : \boldsymbol{p} = \boldsymbol{p}_0 \;\; vs \;\; H_1 : \boldsymbol{p} \neq \boldsymbol{p}_0$。

例子–续

检验统计量为

$$\begin{aligned}\chi^2 &= \frac{(x_1 - np_{01})^2}{np_{01}} + \frac{(x_2 - np_{02})^2}{np_{02}} \\ &\quad + \frac{(x_3 - np_{03})^2}{np_{03}} + \frac{(x_4 - np_{04})^2}{np_{04}} = 0.47.\end{aligned}$$

p 值为 $P(\chi_3^2 > 0.47) = 0.93.$

10.5 置换检验

置换检验的算法

1. 计算检验统计量的观测值

$$t_{\text{obs}} = T(X_1, X_2, \ldots, X_m, Y_1, Y_2, \ldots, Y_n).$$

2. 随机置换数据，用置换数据再次计算检验统计量。
3. 重复前面的过程 B 次，令 $T_1, T_2, \ldots, T_B$ 表示结果值。
4. 近似的 p 值为

$$\frac{1}{B} \sum_{i=1}^{B} I(T_j > t_{\text{obs}}).$$

例子

例 8

令 $T(X_1, X_2, \ldots, X_m, Y_1, \ldots, Y_n) = |\bar{X} - \bar{Y}|$，假设数据为 $(X_1, X_2, Y_1) = (1, 9, 3)$。置换为

置换	T 的值	概率
$(1,9,3)$	2	$1/6$
$(9,1,3)$	2	$1/6$
$(1,3,9)$	7	$1/6$
$(3,1,9)$	7	$1/6$
$(3,9,1)$	5	$1/6$
$(9,3,1)$	5	$1/6$

p 值为 $P(T > 2) = 4/6$.

作业：7

10.6 似然比检验

似然比检验

在参数估计中，我们已经学习了似然函数以及极大似然估计。在假设检验中，也可以基于似然函数得到检验统计量，先看一个简单的例子:

$$H_0 : \ \theta = \theta_0 \text{ v.s. } H_1 : \ \theta \neq \theta_0.$$

对于一组样本 $X_1, \cdots, X_n$,

- 似然函数 $L(\theta)$ 反映了每个参数 θ 产生当前样本的"可能性大小"(似然);
- 因此，$L(\theta_0)$ 和 $\max L(\theta)$ 的接近程度可作为检验标准;

似然比检验

一般的，

定义 6 (似然比检验统计量)

考虑检验

$$H_0: \ \theta \in \Theta_0 \ \ v.s. \ \ H_1: \ \theta \notin \Theta_0.$$

似然比检验统计量为

$$\lambda = 2\log\left(\frac{\sup_{\theta\in\Theta} L(\theta)}{\sup_{\theta\in\Theta_0} L(\theta)}\right) = 2\log\left(\frac{L(\widehat{\theta})}{L(\widehat{\theta}_0)}\right)$$

其中，$\widehat{\theta}$ 是极大似然估计，$\widehat{\theta}_0$ 是 θ 限制在 Θ_0 上的极大似然估计。

基于似然比的思想，可以得到对应的检验统计量:

- 精确分布: 可以通过似然比检验得到简化的检验统计量，然后在原假设下严格推导统计量的分布从而得到拒绝域；
- 大样本渐近分布: 当样本量很大时候 $n \to \infty$, 在原假设下
$$\lambda \rightsquigarrow \chi^2,$$
其中自由度为 Θ 和 Θ_0 的维度之差。

定理 8

假设 $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_q, \theta_{q+1}, \ldots, \theta_r)$，令

$$\boldsymbol{\Theta}_0 = \{\boldsymbol{\theta} : (\theta_{q+1}, \ldots, \theta_r) = (\theta_{0,q+1}, \ldots, \theta_{0,r})\}.$$

令 λ 是似然比检验统计量。在 $H_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ 成立的假设下，

$$\lambda \rightsquigarrow \chi^2_{r-q}.$$

其中 $r-q$ 是 $\boldsymbol{\Theta}$ 的维数减去 $\boldsymbol{\Theta}_0$ 的维数，检验的 p 值为 $P(\chi^2_{r-q} > \lambda)$。

例子

例 9

可以验证，*Mendel* 的豌豆实验中，似然比检验统计量为

$$\begin{aligned}\lambda &= 2\log\left(\frac{L(\widehat{\boldsymbol{p}})}{L(\boldsymbol{p}_0)}\right) \\ &= 2\sum_{j=1}^{4} X_j \log\left(\frac{\widehat{p}_j}{p_{0j}}\right) = 0.48.\end{aligned}$$

p 值为 $P(\chi_3^2 > 0.48) = 0.92$.

例子–续

例 10

假设 $X_1, \cdots, X_n \sim N(\mu, \sigma^2)$, 检验 $H_0: \ \mu = 0, \ \sigma = 1.$
似然比检验统计量为

$$
\begin{aligned}
\lambda &= 2\log\left(\frac{L(\widehat{\mu}, \widehat{\sigma}^2)}{L(0,1)}\right) \\
&= 2\log\left(\frac{L(\bar{X}, \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2)}{L(0,1)}\right) \\
&= \frac{1}{2}(\sum_{i=1}^{n} X_i^2 - n) - \log\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right).
\end{aligned}
$$

作业：13（必做），14（必做），15（选做）

10.7 多重假设检验

10.1 Wald 检验
10.2 p 值
10.3 χ^2 分布
10.4 多项分布数据的 Pearson χ^2 检验
10.5 置换检验
10.6 似然比检验
10.7 多重假设检验
10.8 拟合优度检验

多重假设检验

在统计问题中，有时候我们需要同时做多个假设检验:

$$H_{0i} \quad \text{v.s.} \quad H_{1i}, \; i = 1, \cdots, m.$$

对于每个假设检验，我们都可以得到一个检验统计量 T_i 以及在水平 α 下的拒绝域 $\mathcal{A}_i(\alpha)$. 由检验水平的定义，对于每一个 i,

$$P(T_i \in \mathcal{A}_i(\alpha) | H_{0i}) \leq \alpha.$$

- 对于每个单独的检验，检验水平很好的控制了第一类错误；
- 多个检验时，错误会积累使得无法统一的控制第一类错误。

Bonferroni 方法

一个最简单的方式，我们让每个单独检验的水平为 α/m，这样累计的第一类错误就可以很好的控制在 α 水平内了。

Bonferroni 不等式

$$P(\bigcup_{i=1} A_i) \leq \sum_{i=1} P(A_i).$$

Bonferroni 方法

定义 7 (Bonferroni 方法)

假设 m 个检验的 p 值分别为 $P_1, P_2, \ldots, P_m$，如果

$$P_i < \frac{\alpha}{m},$$

则拒绝原假设 H_{0i}。

Bonferroni 方法的优缺点

- Bonferroni 假设检验方法思想简单，实际中易于操作；
- 缺点是当同时检验个数 m 很大时候, 检验会过于保守, 每一个的拒绝域都会非常小 (接受域很大)，从而倾向于接受原假设.

FDP, FDR

表 1: 多重检验中结果的类型

	不拒绝 H_0	拒绝 H_0	总计
H_0 为真	U	V	m_0
H_0 为假	T	S	m_1
总计	$m-R$	R	m

错误发现比例 (FDP) 为

$$\mathrm{FDP} = \begin{cases} V/R & R > 0 \\ 0 & R = 0. \end{cases} \qquad \mathrm{FDR} = \mathrm{E}(\mathrm{FDP}).$$

BH 方法

定义 8 (BH 方法)

1. 令 $P_{(1)} < P_{(2)} < \cdots < P_{(m)}$ 表示排序后的 p 值;
2. 定义 $\ell_i = \dfrac{i\alpha}{mC_m}$, $R = \max\{i : P_{(i)} < \ell_i\}$, 其中，如果 p 值独立，则 C_m 定义为 *1*，否则令 $C_m = \sum_{i=1}^{m}(1/i)$。
3. 令 $T = P_{(R)}$，则称 T 为 *BH* 拒绝阀。
4. 拒绝所有 $P_i \le T$ 的原假设 H_{0i}。

定理 9

BH 方法满足 $FDR \le \frac{m_0}{m}\alpha \le \alpha$。

例 11

10 个相互独立的假设检验的 p 值分别为：

$$0.00017, 0.00448, 0.00671, 0.00907, 0.01220,$$
$$0.33626, 0.39341, 0.53882, 0.58125, 0.98617.$$

给定 $\alpha = 0.05$，

(1) Bonferroni 拒绝 p 值小于 $0.05/10 = 0.005$ 的检验；

(2) BH:$\ell_i = i * 0.05/10 = 0.005i$，

$$R = \max\{i : p_{(i)} < 0.005i\} = 5,$$

拒绝 p 值小于或等于 $p_{(5)} = 0.01220$ 的检验。

作业：11（必做），12（选做）

10.8 拟合优度检验

令 $\mathcal{F} = \{f(x;\boldsymbol{\theta} : \boldsymbol{\theta} \in \boldsymbol{\Theta})\}$ 为参数模型，假设数据在实数线上取值，把实数线分成 k 个不相交的区间 $I_1, I_2, \ldots, I_k$。定义

$$p_j(\boldsymbol{\theta}) = \int_{I_j} f(x;\boldsymbol{\theta})dx, \quad j = 1, 2, \ldots, k.$$

令 N_j 为落入 I_j 的观测数，基于 $N_1, N_2, \ldots, N_k$ 的 θ 的似然函数为多项分布似然函数

$$Q(\boldsymbol{\theta}) = \prod_{j=1}^{k} p_j(\boldsymbol{\theta})^{N_j}.$$

$Q(\boldsymbol{\theta})$ 最大化得到的 $\boldsymbol{\theta}$ 的估计为 $\tilde{\boldsymbol{\theta}}$，定义检验统计量为

$$Q = \sum_{j=1}^{k} \frac{(N_j - np_j(\tilde{\boldsymbol{\theta}}))^2}{Np_j(\tilde{\boldsymbol{\theta}})}.$$

$H_0: \boldsymbol{\theta}$ 为模型的真实参数，则

定理 10

$Q \rightsquigarrow \chi^2(k-1-s)$，$s$ 为 Θ 的维数。近似的 p 值为 $p = P(\chi^2(k-1-s) > q)$，q 表示 Q 的观测值。