论假设检验中的两类错误*

蔡越江

(北京工业大学计算机学院, 北京, 100022)

摘 要

蔡越江. 论假设检验中的两类错误.

本文从假设检验的两个例子谈起, 指出备择假设也是影响接区域的。接着论述了假设检验中的两类错误之间的关系。最后讨论如何同时控制两类错误。

关键词: 假设检验 两类错误 功效函数

一、引言

数理统计学中的假设检验方法已在许多领域得到广泛应用, 特别在科学实验与质量检验中。但在应用中也出现一些问题。为使这一方法能更有效地应用于实际工作, 我们通过纠正一篇文章中的模型, 来说明如何正确地使用假设检验方法。重点是论述假设检验中的两类错误。

二、假设检验的两个例子

本文以文献[1]的术语和符号为准。设有样本 X, 取值于样本空间 $\mathscr{X}$。只知道 X 的分布属于一个以 θ 为参数的分布族 $\{F_\theta, \theta\in\Theta\}$。设 Θ_H 是 Θ 的一个非空真子集, 则命题 $H_0: \theta\in\Theta_H$ 称为原假设。命题 H_0 的确切含义是: 存在一个 $\theta_0\in\Theta_H$ 使样本 X 的分布为 F_{θ_0}。记 $\Theta_K=\Theta-\Theta_H$, 则命题 $H_1: \theta\in\Theta_K$ 称为 H_0 的对立假设, 也有称为备择假设。表述 $H_0: \theta\in\Theta_H \leftrightarrow H_1: \theta\in\Theta_K$ 称为一个假设检验问题。其意义是: 根据 X 的具体值判断 H_0 是否正确, 或者说在 H_0, H_1 中选择一个, 分别称为接受 H_0 和否定 H_0。在接验一个假设时, 所使用的统计量称为检验统计量。使原假设 H_0 得到接受的那些样本 X 所在的区域 A 称为该检验的接受区域, 而使原假设被否定的那些样本所在的区域称为否定区域。

例1 将一枚硬币投掷20次, 以正面出现次数的多少来断言硬币是否完好。

我们知道, 若硬币是完好的, 则正面出现的概率 $P=0.5$, 可见投掷20次正面出现的平均值为10次。考虑到硬币(正面与反面)和二项分布的对称性, 我们认为正面出现的次数特多或特少的硬币是有制陷的。我们作如下假设:

原假设 $H_0: P=0.5$; 备择假设 $H_1: P\neq 0.5$ 选择检验统计量 $X=\sum_{i=1}^{20} x_i$ (下标 i 标志第 i 次投掷, 出现正面 $x_i=1$; 出现反面 $x_i=0$), 经过计算可知投掷20次, 正面出现6～14次之间的概

* 收稿日期: 1999年1月28日

率为 0.9556。当置信水平 $\alpha=1-0.9556=0.0444$ 时，我们可选取对称的接受区域 $A=\{6\le\sum_{i=1}^{20}x_i\le14\}$。这时，犯第一类错误(弃真)的概率 $\alpha=0.0444$。

例 2　某刊物出现这样的假设检验模型：原假设 H_0：$P=0.5$；备择假设 H_1：$P=0.8$(投掷时出现正面的机会很多)。求当 H_1 为真时错误接受 H_0 的概率，即求在 H_0 下，$\alpha=0.0444$，在 $1-\alpha=0.9556$ 的接受区域$\{6\le\sum_{i=1}^{20}x_i\le14\}$中所包含的 H_1 分布的面积，这个面积就是第二类错误的概率 β。* 经过计算得 $\beta=0.2$。计算时使用了平均值 $\mu=20\times0.8=16$，标准差 $\sigma=\sqrt{20\times0.8\times(1-0.8)}$，即两处使用了 $P=0.8$ 这个数据。虽然 $\beta=0.2$ 是计算出来了，但这个 0.2 是不合理的，因为例二的接受区域{出现正面的次数在 6～14 之间}是不合适的。确定接受区域既要考虑原假设 H_0，同时要考虑备择假设 H_1。不考虑 H_1 将导致一些错误的结果。我们考虑掷币 20 次，正面出现的次数为 3 次时，它落在接受区域 A 之外，即落在否定区域内。这时我们就否定原假设 H_0：$P=0.5$。我们就要去接受这个准备供选择的备择假设 H_1：$P=0.8$(正面出现的概率 0.8)。掷币 20 次正面只出现 3 次，要去否定正面出现的概率 0.5 而去接受正面出现的概率 0.8，这是明显的错误。这错误是由于接受区域 A 不合适引起的。有人计算过，在 $P=0.5$ 时，正面出现次数为 3 的概率为 0.0011，它是 $P=0.8$ 条件下出现 3 次正面的概率 7.65E(-10)的 142 万倍。此时不接受原有假设，而去接受备择假设能合理吗？所以在选择判别区域时，不能只考虑原假设，还要认真考虑备择假设，它们是不可分割的整体，只有这样，才能选择出合理有效的判别区域。

对原假设 H_0：$P=0.5$；备择假设 H_1：$P=0.8$ 这个问题来说，这时接受区域应该是{出现正面的次数在 0～C 之间}即$\{\sum_{i=1}^{20}\le C\}$，然后根据置信水平确定合适的临界值 C。这时$\{\sum_{i=1}^{20}\le C\}$中所包含的 H_1 分布的面积就是第二类错误的概率 β 的数值。这个数值 β 是容易计算的。

有些人强调当接受 H_0 时，应该知道这个决策出错的概率 β，但例一这个模型是不可能计算第二类错误 β 的数值。那么这个要求在通常的假设检验中能否实现呢？为此我们通过实例讨论一下第一与第二类错误之间的关系，以及统计中对这些问题的处理。

三、两类错误间的关系

犯第二类错误 β 数值的大小，是同根据置信水平 α 确定的接受区域的位置有关(不管是双侧检验还是单侧检验)。同时均值 μ_0 与 μ_1 的距离对 β 的大小影响很大。[2]。即在假设检验问题中，仅关心 H_0 是不够的，必须同时注意备择假设 H_1，H_1 同样影响 β 数值的大小。如果 β 数值很大，比如 $\beta=0.5$，即有二分之一的可能犯第二类错误，这样的统计结论的意义就不大了。

对初学假设检验者来说，不少人(包括我本人)往往有一个错觉，认为犯第一类错误 α 的数值越小越好，这是不正确的。一般来说，当其它条件不变时，α 小，则 β 增大；反之 α 大，必导致 β 减小。它表明其它条件不变的情况下，特别是样本数 N 确定后，要 α 与 β 都很小是不可能的。

这里顺便提一下：在规定的 α 数值下，将单侧检验与双侧检验相比较，单侧检验将得到较小的 β。[2](P. 129)因此，在实际工作中，凡是可以采用单侧检验的场合，就应尽量利用这一有利条件，以有效地减小 β 的数值。我们在上节中和文[3]结束语末尾指出的接受区域$\{\sum_{i=1}^{20}x_i\le C\}$，用的就是单侧检验。再根据 Neyman-Pearson 引理，还可知我们选择的接受区域是一致最

* H_1 为真则 H_0 为伪，接受 H_0 的概率即采伪错误的概率。

优检验,即在限制第一类错误的概率不超过 α 的条件下,总使第二类错误的概率 β 达到最小。有兴趣的同志可参阅任何一本数学系数理统计专业使用的数理统计学教材。

我们知道,通常利用假设检验解决问题时,只考虑并控制犯第一类错误的概率,那么为什么不计算、不限制犯第二类错误的概率呢? 主要原因有两条:

第一 第一类错误和第二类错误之间存在密切关系。前已指出,当样本容量 n 确定后,不能同时控制犯第一类错误的概率 α 和第二类错误的概率 β 都很小。例如,我们考虑一种极端情况,即不犯第一类错误(弃真错误),这时 $\alpha=0$ 接受区域是全部样本空间,而拒绝区域则为空集。这时,犯第二类错误的概率 $\beta=P_\theta\{X\in\mathscr{X}\}=1$。同样,如果不犯第二类错误(取伪错误),$\beta=0$,这时拒绝区域就是全部样本空间,此时 $\alpha=1$。一般来说,当其它条件不变时,α 大,则 β 小;反之 α 小,必导致 β 增大。所以假设检验不是 α 越小越好,α 小隐含着 β 越大。在实际应用时,必须根据客观事物的背景选取合适的 α 或合适的 β。所以通常人们选取接受区域为犯第一类错误的概率不超过 α 的前提下,犯第二类错误的概率 β 尽量小。

至于 α 的选取,要看问题的实际背景和使用者的企图。"通常人们不希望轻易拒绝 H_0,例如工厂的产品一般是合格的,出厂进行抽样检查时不希望轻易地被认为不合格,于是在限定犯第一类错误的概率不超过某个指定值 α 的条件下,寻求犯第二类错误的概率尽可能小的检验方法"(中国大百科全书,数学,P360)。这时,为保护原假设 H_0,α 可以取小一点(最常用的是 $\alpha=0.05$ 和 0.01,有时也用到 0.001, 0.10),β 值往往控制在 0.10～0.30。陈希孺院士在[4]中指出:"从实用的观点看,确实,在多数假设检验问题中,第一类错误被认为更有害,更需要控制。""也有些情况,确实第二类错误的危害更大,这时有必要控制这个概率,换句话说'控制第一类错误'的原则也并非绝对的,可视情况的需要而变通之。"(P. 222)

第二:第二类错误的概率通常很给计算或根本计算不出来。

举一个最简单的例子,上节中针对判定硬币是否完好的假设检验:原假设 H_0:$P=0.5$;备择假设 H_1:$P\neq0.5$,这时由于备择假设是两个连续的区间$[0,0.5)$、$(0.5,1]$。其犯第二类错误的概率 β 根据不同的备择假设 P 其范围是$(0,1-\alpha)$。这样简单的模型都不能计算出具体的 β 值,当然也就无法控制这个 β 值。更不用说复杂一些的模型了。

对例二来说,在原假设 H_0:$P=0.5$;备择假设 H_1:$P=0.8$ 条件下,根据置信水平 α,确定临界值 C,再根据$\{\sum_{i=1}^{20}x_i\leq C\}$计算出犯第二类错误的概率 β。而这个假设对判定硬币是否有缺陷是没有多大用处的,它只能对出现正面的概率是 0.5,还是 0.8 作出判定。所以在这种情况下,即使第二类错误 β 的数值能计算出来,但对判定硬币是否完好却没有实际意义。

考虑第二类错误 β 是非常正确的,也是很有必要的,但是真正计算出有实际使用价值的 β 是很困难的,有时甚至是不可能的。

四、如何同时控制两类错误

在一些实际问题中,人们不仅希望控制犯第一类错误的概率,而且希望适当地控制犯第二类错误的概率。当样本个数 n 确定后,不能同时控制犯两类错误的概率都很小。因此,只能通过加大要本容量 n,达到同时控制两类错误的目的。那么,样本容量 n 增加到多少合适呢? n 太大,则费用增加太大,太小又达不到精度。因而,通常采用的方法是预先给出显著水平 α,控制犯第一类错误的概率。再选取样本容量 n,将犯第二类错误的概率控制在预先给定的限度之

内。

为了控制犯第二类错误的概率, 首先引入功效函数这一重要概念。

设有样本 X 取值于样本空间 $\mathscr{X}$, 并且知道 X 的分布属于一个分布族 $\{F_\theta: \theta\in\Theta\}$, 设假设检验问题为: 原假设 $H_0: \theta\in\Theta_0$ ↔ 备择假设 $H_1: \theta\in\Theta_1$ $\Theta_0\cap\Theta_1=\phi$, 并且 Θ_0 和 Θ_1 是 Θ 的非空真子集。这个假设检验问题的一个检验 φ_n 为: 将样本空间 $\mathscr{X}$ 分解为两个互不相交的部分 $\mathscr{X}_1$ 和 $\mathscr{X}_2$, $\mathscr{X}_1$ 为接受区域, $\mathscr{X}_2$ 为否定区域, 这个检验 φ_n 的功效函数定义为:

$$\beta_n(\theta) = P_\theta(\text{否定原假设}) = P_\theta(X\in\mathscr{X}_2)$$

知道了一个检验的功效函数 $\beta_n(\theta)$, 就可算出它犯各类错误的概率:

犯第一类错误的概率 $\alpha=\beta_n(\theta)$ $\theta\in\Theta_0$ 犯第二类错误的概率 $\beta=1-\beta_n(\theta)$ $\theta\in\Theta_1$

如果功效函数 $\beta_n(\theta)$ 满足下述条件: $\lim\limits_{n\to\infty}(1-\beta_n(\theta))=0$ 对 $\theta\in\Theta_1$ 一致成立。则可以选择合适的 n, 控制犯第二类错误的概率。

下面针对前面提到的判别硬币是否完好的假设检验问题, 说明如何同时控制两类错误。

原假设 $H_0: \theta=0.5$ ↔ 备择假设 $H_1: \theta\neq 0.5$ 根据样本容量 n 确定的检验 φ_n 为:

接受区域 $\mathscr{X}_1=\left\{\sqrt{n}\left|\dfrac{\overline{X}-0.5}{0.5}\right|\leq z_{\frac{\alpha}{2}}\right\}$ 拒绝区域 $\mathscr{X}_2=\left\{\sqrt{n}\left|\dfrac{\overline{X}-0.5}{0.5}\right|> z_{\frac{\alpha}{2}}\right\}$

其中 $\overline{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i$ 是样本均值, $\Phi(x)$ 表示标准正态分布函数, z_α 为满足 $\Phi(z_\alpha)=1-\alpha$ 的分位点。

由于样本容量 n 较大($n\geq 20$), 我们采用正态分布近似估计检验 φ_n 的功效函数。因为样本均值 $\overline{X}$ 的数学期望为 θ, 方差为: $\dfrac{\theta(1-\theta)}{n}$, 所以它的功效函数为: $\beta_n(\theta)=P_\theta(X\in\mathscr{X}_2)$

犯第一类错误的概率 $\beta_n(0.5)=\alpha$

犯第二类错误的概率:

$$1-\beta_n(\theta)=P_\theta\left\{\sqrt{n}\left|\frac{\overline{X}-0.5}{0.5}\right|\leq z_{\frac{\alpha}{2}}\right\} \qquad \theta\in\Theta_1 \text{ 即 } \theta\neq 0.5$$

$$=P_\theta\left\{\frac{-\frac{0.5}{\sqrt{n}}z_{\frac{\alpha}{2}}+0.5-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\leq\frac{\overline{X}-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\leq\frac{\frac{0.5}{\sqrt{n}}z_{\frac{\alpha}{2}}+0.5-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\right\}$$

$$=\Phi\left(\frac{\frac{0.5}{\sqrt{n}}z_{\frac{\alpha}{2}}+0.5-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\right)-\Phi\left(\frac{-\frac{0.5}{\sqrt{n}}z_{\frac{\alpha}{2}}+0.5-\theta}{\sqrt{\frac{\theta(1-\theta)}{n}}}\right) \qquad (1)$$

$$=\Phi\left(\frac{0.5}{\sqrt{\theta(1-\theta)}}z_{\frac{\alpha}{2}}+\sqrt{n}\frac{0.5-\theta}{\sqrt{\theta(1-\theta)}}\right)-\Phi\left(\frac{-0.5}{\sqrt{\theta(1-\theta)}}z_{\frac{\alpha}{2}}+\sqrt{n}\frac{0.5-\theta}{\sqrt{\theta(1-\theta)}}\right) \qquad (2)$$

依上式, 得 $\lim\limits_{n\to\infty}(1-\beta_n(\theta))$ 不一致趋于 0, 故找不到公共的 n_0, 当 $n\geq n_0$ 时 $0<1-\beta_n(\theta)<\beta$。

现在, 我们重新考虑一下问题的背景, 即判定一枚硬币是否完好。所以当出现正面的概率 θ 与 0.5 很接近时, 通常就认为硬币是完好的。所以要求控制 H_1 中 $|\theta-0.5|\geq\delta$ 的 θ, 犯第二类错误的概率小于或等于 β。具体办法如下: 首先考虑 $0<\theta\leq 0.5-\delta$ 的情况。

$$\because \lim_{n\to\infty}\frac{0.5}{n}z_{\frac{\alpha}{2}}=0,\ \therefore \exists\ n_1 \quad 当\ n\ge n_1\ 时\ \frac{0.5}{n}z_{\frac{\alpha}{2}}<\frac{\delta}{2} \quad 即:\ \begin{aligned}-\frac{0.5}{n}z_{\frac{\alpha}{2}}+0.5-\theta\ge\frac{\delta}{2}\\ \frac{0.5}{n}z_{\frac{\alpha}{2}}+0.5-\theta\le\frac{3}{2}\delta\end{aligned} \tag{3}$$

由(1)式和(3)式知: 当 $n>n_1$ 时

$$0<1-\beta_n(\theta)\le\Phi\left(\frac{3}{2}\delta\sqrt{\frac{n}{\theta(1-\theta)}}\right)-\Phi\left(\frac{\delta}{2}\sqrt{\frac{n}{\theta(1-\theta)}}\right)\xrightarrow{\theta\to 0}0 \tag{4}$$

对任意 n 一致成立。

$\therefore\exists\ \theta_1$ 当 $0<\theta<\theta_1$, $n>n_1$ 时 $0<1-\beta_n(\theta)\le\beta$ (5)

当 $\theta_1\le\theta\le 0.5-\delta$ 时

由 $\frac{0.5z_{\frac{\alpha}{2}}}{\sqrt{\theta(1-\theta)}}\le\frac{0.5z_{\frac{\alpha}{2}}}{\sqrt{\theta_1(1-\theta_1)}}$ 和 $\frac{0.5-\theta}{\sqrt{\theta(1-\theta)}}\ge\frac{0.5-\theta}{0.5}\ge 2\delta>0$ 知:

$$\begin{aligned}&\frac{0.5}{\sqrt{\theta(1-\theta)}}z_{\frac{\alpha}{2}}+\sqrt{n}\,\frac{0.5-\theta}{\sqrt{\theta(1-\theta)}}\xrightarrow{n\to\infty}+\infty\\ -&\frac{0.5}{\sqrt{\theta(1-\theta)}}z_{\frac{\alpha}{2}}+\sqrt{n}\,\frac{0.5-\theta}{\sqrt{\theta(1-\theta)}}\xrightarrow{n\to\infty}+\infty\end{aligned} \tag{6}$$

对任意 $\theta_1\le\theta\le 0.5-\delta$ (6)式一致成立。

根据(2)式和(6)式知 $\lim\limits_{n\to\infty}(1-\beta_n(\theta))=0$ 对 $\theta_1\le\theta\le 0.5-\delta$ 一致成立,即∃ n_2 当 $n>n_2$ 时

$$0<1-\beta_n(\theta)<\beta \tag{7}$$

由(5)式和(7)式知,当 $n>n_0=\max(n_1,n_2)$ 时,对任意满足 $0\le\theta\le 0.5-\delta$ 的 θ 有

$$0<1-\beta_n(\theta)<\beta \quad 成立$$

同样的方法可以找到 n_3 当 $n>n_3$ 是,对任意满足 $0.5+\delta\le\theta<1$ 的 θ 有 $0\le 1-\beta_n(\theta)\le\beta$ 成立。这样通过对功效函数的计算,就能够选择合适的 n^*,当 $n\ge n^*$ 时犯第二类错误的概率不超过预先给定的 β。

五、结束语

假设检验是一种有重要应用价值的统计推断形式,在实际工作中其应用日趋普遍、日趋深刻。由于往往有人只根据显著性水平 α 选择置信区间,没有深入考虑犯第二类错误的概率;也有人认为第一类错误与第二类错误同等重要,所以一定要计算出犯第二类错误的概率;也有些人将假设检验中的原假设与备择假设割裂开来,视备择假设为摆设。通过本文的论述,使大家对假设检验方法有更深的了解,同时提供一种控制犯第二类错误概率的思路供大家参考。

参考文献

[1] 陈希孺,倪国熙,数理统计学教程,上海;上海科学技术出版社,1988
[2] 潘维栋,数理统计方法,上海教育出版社,1980
[3] 蔡越虹,蔡越江。再论《统计假设检验在地学应用中的两种"错误"》,物探化探计算技术,第 20 卷第 2 期,1998
[4] 陈希孺,概率论与数理统计,合肥:中国科学技术大学出版社,1992

(下转封三)

(上接第 34 页)

Note On Two Types Of Errors Of Statistical Hypothesis Tests

Cai Yuejiang
(Computer College of Beijing Polytechnic University)

Abstract

In this paper, I started from two examples of hypothesis tests and pointed out that alter-native hypothesis influence the accept area, and then I talked about the realtion between two types of errors in statistical hypothesis. At last I discussed about how to control the two types of errors in the same time.

Key Words: hypothesis test, two types of errors, efficacy function.

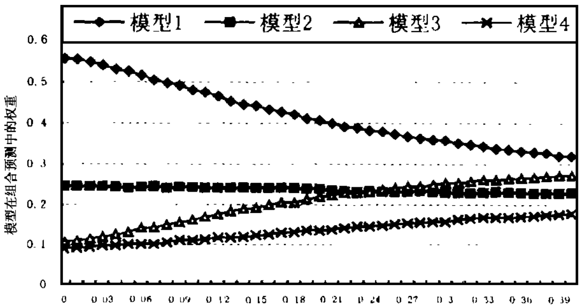

图 5　4 种模型近 8 年整体权重平均值随 α (横坐标) 变化的曲线