假设检验

December 9, 2017

假设检验

[1] 在总体的分布函数完全未知或只知其形式、但不知其参数的情况，为了推断总体的某些未知特性，提出某些关于总体的假设。根据样本对所提出的假设作出是接受，还是拒绝的决策。

提出两个相互独立的假设

$$H_0 :=$$

$$H_1 :\neq \quad .$$

然后给出一个合理的法则，根据这一法则，利用已知样本作出决策是接受假设 H_0(拒绝假设 H_1)，还是拒绝假设 H_0(接受假设 H_1)。由于样本均值 $\overline{X}$ 是 μ 的无偏估计，$\overline{X}$ 的观察值 $\overline{x}$ 的大小在一定程度上反映 μ 的大小。若假设 H_0 为真，则观察值 $\overline{x}$ 与 μ_0 的偏差 $|\overline{x}-\mu_0|$ 一般不应太大。若 $|\overline{x}-\mu_0|$ 过分大，就怀疑假设 H_0 的正确性而拒绝 H_0，并考虑到当 H_0 为真时 $\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$。衡量 $|\overline{x}-\mu_0|$ 的大小可归结为衡量 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}}$ 的大小。适当选定一正数 k，使当观察值 $\overline{x}$ 满足 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} \geqslant k$ 时拒绝假设 H_0，反之，若 $\dfrac{|\overline{x}-\mu_0|}{\sigma/\sqrt{n}} < k$，就接受假设 H_0。

然而，由于作出决策的依据是一个样本，当实际上 H_0 为真时仍可能作出拒绝 H_0 的决策，这种可能性是无法消除的。这是一种错误，犯这种错误的概率记为

$P\{$ 当 H_0 为真拒绝 $H_0\}$ 或 $P_{\mu_0}\{$ 拒绝 $H_0\}$ 或 $P_{\mu\in H_0}\{$ 拒绝 $H_0\}$

$P_{\mu_0}\{\cdot\}$ 表示参数 μ 取 μ_0 时事件 $\{\cdot\}$ 的概率，$P_{\mu\in H_0}\{\cdot\}$ 表示 μ 取 H_0 规定的值时事件 $\{\cdot\}$ 的概率。希望将犯这类错误的概率控制在一定限度内，即给出一个较小的整数 $\alpha(0<\alpha<1)$，使犯这类错误的概率不超过 α，即使得

$$P\{\text{当}H_0\text{为真拒绝}H_0\}=P_{\mu_0}\left\{\left|\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant k\right\}\leqslant\alpha\ . \tag{1}$$

当 H_0 为真时，$Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\sim N(0,1)$，由标准正态分布分位点的定义

$$k=z_{\alpha/2}\ . \tag{2}$$

若Z 的观察值满足

$$|z|=\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant k=z_{\alpha/2}\ ,$$

则拒绝 H_0，若

$$|z|=\left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|<k=z_{\alpha/2}\ ,$$

则接受 H_0。通常 α 总是取得较小，$\alpha=0.01,0.005$。若 H_0 为真，即当 $\mu=\mu_0$ 时，$\left\{\left|\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant z_{\alpha/2}\right\}$ 是一个小概率事件，根据实际推断原理可以认为，若 H_0 为真，则由一次试验得到的观察值 $\overline{x}$ 满足不等式 $\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant z_{\alpha/2}$ 几乎是不会发生的。如果在一次观察中出现了满足 $\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant z_{\alpha/2}$ 的 $\overline{x}$，则有理由怀疑原先的假设 H_0 的正确性，因而拒绝 H_0。若出现的观察值 $\overline{x}$ 满足 $\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|<z_{\alpha/2}$，此时没有理由拒绝假设 H_0，即接受假设 H_0。

当样本容量固定时，选定 α 后，数 k 就可以确定，然后按照统计量 $Z=\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 的观察值的绝对值 $|z|$ 大于等于 k 还是小于 k 作出决策。若 $|z|=\left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right|\geqslant k$，则称 $\overline{x}$ 与

μ_0 的差异是显著的，拒绝 H_0；若 $|z| = \left|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| < k$，则称 $\overline{x}$ 与 μ_0 的差异是不显著的，接受 H_0。α 称为显著性水平。统计量 $Z = \dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 称为检验统计量。

检验问题通常叙述成：在显著性水平 α 下，检验假设

$$H_0 : \mu = \mu_0\ ,\quad H_1 : \mu \neq \mu_0\ . \tag{3}$$

也即"在显著性水平 α 下，针对 H_1 检验 H_0"。H_0称为原假设或零假设，H_1称为备择假设(在原假设被拒绝后可供选择的假设)。

当检验统计量取某个区域 C 中的值时，拒绝原假设 H_0，称区域 C 为拒绝域，拒绝域的边界点称为临界点。以上问题的拒绝域为 $|z| \geqslant z_{\alpha/2}$，而 $z = -z_{\alpha/2}, z = z_{\alpha/2}$ 为临界点。在假设 H_0 实际为真时，可能犯拒绝 H_0 的错误，称为"弃真"错误，也叫第 I 类错误。当 H_0 实际为不真时，有可能接受 H_0，称为"取伪"错误，也叫第 II 类错误。

犯第 II 类错误的概率

$P\{$ 当 H_0 不真接受 $H_0\}$ 或 $P_{\mu\in H_1}\{$ 接受 $H_0\}$

在确定检验法则时，尽可能使犯两类错误的概率都较小。但当样本容量固定时，若减少犯一类错误的概率，则犯另一类错误的概率往往增大。要使犯两类错误的概率都减小，除非增加样本容量。在给定样本容量的情况下，总是控制犯第 | 类错误的概率，使它不大于 α。只对犯第 I 类错误的概率加以控制，而不考虑犯第 II 类错误的概率的检验，称为显著性检验。若备择假设 H_1 中，μ 可能大于 μ_0，也可能小于 μ_0，称为双边备择假设，该假设检验为双边假设检验。

$$H_0 : \mu \leqslant \mu_0, \quad H_1 : \mu > \mu_0$$

称为右边检验，

$$H_0 : \mu \geqslant \mu_0, \quad H_1 : \mu < \mu_0$$

称为左边检验。右边检验和左边检验统称单边检验。

处理参数的假设检验步骤如下:

1. 根据实际问题的要求，提出原假设 H_0 及备择假设 H_1;
2. 给定显著性水平 α，及样本容量 n;
3. 确定检验统计量及拒绝域的形式;
4. 按 $P\{$ 当 H_0 为真拒绝 $H_0\} \leqslant \alpha$ 求出拒绝域;
5. 取样，根据样本观察值作出决策，是接受还是拒绝 H_0。

Decide between two competing statements, H_0 and H_a, called the null hypothesis and alternative hypothesis, based on the data. There are two possible errors in adjudicating between these hypotheses:

Type 1 error Here one wrongly rejects the null hypothesis H_0 giving a false positive decision.

Type 2 error Here one fails to reject the null hypothesis when the alternative is true, giving a false negative decision. In our example, we would incorrectly infer that a signal is absent when it truly is present.

[实验的数学处理] 选择一个统计量

$$\lambda = \lambda(\boldsymbol{x}) \tag{4}$$

叫做检验统计量，在假设 H 成立的条件下，检验统计量 λ 的概率 (密度) 函数 $p(\lambda|H)$ 已知。由 $p(\lambda|H)$ 可以决定 λ 值的某个区域 ω，叫做拒绝域。整个区域的概率含量

$$P_r(\lambda \in \omega|H) = \int_{\lambda\in\omega} p(\lambda|H)\mathrm{d}\lambda = \alpha \tag{5}$$

也很小 ($\lambda \in \omega$ 表示 λ 值属于区域 ω)。若假设 H 为真，则 λ 值落入拒绝域 ω 内的可能性很小。若由样本 $\boldsymbol{x}$ 所算得的统计量 λ 的数值 $\lambda = \lambda(\boldsymbol{x})$ 落到了拒绝域 ω 内，则说观

测结果 $\boldsymbol{x}$ 同假设 H 有显著的矛盾 (显著水平 α)，或者说在显著水平 α 下拒绝假设 H。若统计量 $\lambda = \lambda(\boldsymbol{x})$ 落在区域 ω 外，则样本 $\boldsymbol{x}$ 同假设 H 没有显著的矛盾 (显著水平 α)，或者说在显著水平 α 下接受假设 H。

显著水平 α 就是在原假设 H 成立的条件下，检验统计量 λ 在拒绝域 ω 内的概率含量，即分布 $p(\lambda|H)$ 在拒绝域 ω 内的概率含量。

即使原假设 H 为真，λ 值也可能落入拒绝域 ω 内。此时会错误地拒绝原假设 H，显著水平 α 就是犯这种错误地概率。α 是在 H 为真的条件下，λ 落入区域 ω 内的概率。α 是区域 ω 包含 λ 的置信水平。记 λ 的全部可能取值

正态总体均值的假设检验

正态总体方差的假设检验

置信区间与假设检验之间的关系

样本容量的选取

分布拟合检验

单个分布的 χ^2 拟合检验法

设总体 X 的分布未知，$x_1, x_2, \cdots, x_n$ 是来自 X 的样本值。检验假设

H_0：总体 X 的分布函数为 $F(x)$；

H_1：总体 X 的分布函数不是 $F(x)$；

设 $F(x)$ 不含未知参数。(也常以分布律或概率密度代替 $F(x)$)

将在 H_0 下 X 可能取值的全体 Ω 分成互不相交的子集 $A_1, A_2, \cdots, A_k$，以 $f_i(i = 1, 2, \cdots, k)$ 记样本观察值 $x_1, x_2, \cdots, x_n$ 落在 A_i 的个数，表示事件 $A_i = \{X$ 的值落在子集 A_i 内 $\}$ 在 n 次独立试验中发生 f_i 次，在这 n 次试验中事件 A_i 发生的频率为 f_i/n。当 H_0 为真时，根据 H_0 中所假设的 X 的分布函数来计算事件 A_i 的概率，得到 $p_i = P(A_i), i = 1, 2, \cdots, k$。当 H_0 为真，且试验次数很多时，频率 f_i/n 和概率 p_i 的差异不应太大。用统计量

$$\sum_{i=1}^{k} C_i \left(\frac{f_i}{n} - p_i \right)^2 \tag{6}$$

来度量样本与 H_0 中所假设的分布的吻合程度，其中 $C_i(i = 1, 2, \cdots, k)$ 为给定的常数。若选取 $C_i = n/p_i, (i = 1, 2, \cdots, k)$，检验统计量为

$$\chi^2 = \sum_{i=1}^{k} \frac{n}{p_i} \left(\frac{f_i}{n} - p_i \right)^2 = \sum_{i=1}^{n} \frac{f_i^2}{np_i} - n \ , \tag{7}$$

当 n 充分大 $(n \geqslant 50)$，则当 H_0 为真时，该统计量近似服从 $\chi^2(k-1)$ 分布。当 H_0 为真时，χ^2 不应该太大，若 χ^2 过大就拒绝 H_0，拒绝域形式为

$$\chi^2 \geqslant G \tag{8}$$

G 为正常数。对于给定的显著性水平 α，确定 G 使

$P\{$ 当 H_0 为真时拒绝 $H_0\} = P_{H_0}\{\chi^2 \geqslant G\} = \alpha$

得到 $G = \chi^2_\alpha(k-1)$。即当样本观察值使得

$$\chi^2 \geqslant \chi^2_\alpha(k-1) \tag{9}$$

则在显著性水平 α 下拒绝 H_0，否则接受 H_0。n 不能小于 50，$np_i \geqslant 5$，否则适当合并 A_i。

分布族的 χ^2 拟合检验法

原假设

H_0：总体 X 的分布函数是 $F(x;\theta_1,\theta_2,\cdots,\theta_r)$，

F 的形式已知，而 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_r)$ 是未知参数，它们在某一个范围取值。将在 H_0 下 X 可能取值的全体 Ω 分成 $k(k>r+1)$ 个互不相交的子集 $A_1,A_2,\cdots,A_k$，以 $f_i(i=1,2,\cdot,k)$ 记样本观察值 $x_1,x_2,\cdots,x_n$ 落在 A_i 的个数，则事件 $A_i=\{X$ 的值落在 A_i 内 $\}$ 的频率为 f_i/n。当 H_0 为真时，由 H_0 所假设的分布函数来计算 $P(A_i)$，得到 $P(A_i)=p_i(\theta_1,\theta_2,\cdots,\theta_r)=p_i(\boldsymbol{\theta})=p_i$。先利用样本求出未知参数的最大似然估计 (在 H_0 下)，以估计值作为参数值，求出 p_i 的估计值 $\hat{p}_i=\hat{P}(A_i)$。

$$\chi^2=\sum_{i=1}^{k}\frac{f_i^2}{n\hat{p}_i}-n \tag{10}$$

作为检验假设 H_0 的统计量。在某些条件下，在 H_0 为真时，近似地有

$$\chi^2=\sum_{i=1}^{k}\frac{f_i^2}{n\hat{p}_i}-n\sim\chi^2(k-r-1)\ . \tag{11}$$

拒绝域为

$$\chi^2\geqslant\chi^2_\alpha(k-r-1)\ , \tag{12}$$

α 为显著性水平。

偏度、峰度检验

χ^2 拟合检验法用来检验总体的正态性时，犯 || 类错误的概率往往较大。

随机变量 X 的偏度和 峰度是 X 的标准变化量 $[X-E(X)]/\sqrt{D(X)}$的三阶矩和四阶矩：

$$\nu_1 = E\left[\left(\frac{X-E(X)}{\sqrt{D(X)}}\right)^3\right] = \frac{E[(X-E(X))^3]}{(D(X))^{3/2}} ,$$

$$\nu_1 = E\left[\left(\frac{X-E(X)}{\sqrt{D(X)}}\right)^4\right] = \frac{E[(X-E(X))^4]}{(D(X))^2} .$$

当随机变量 X 服从正态分布时，$\nu_1 = 0$ 且 $\nu_2 = 3$。

设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，则 ν_1, ν_2 的矩估计量分别是

$$G_1 = B_3/B_2^{3/2} \ , G_2 = B_4/B_2^2 \ , \tag{13}$$

其中 $B_k(k=2,3,4)$ 是样本 k 阶中心矩，分布称 G_1, G_2 为样本偏度和样本峰度。

若总体 X 为正态分布，当 n 充分大时，近似有

$$G_1 \sim N\left(0, \frac{6(n-2)}{(n+1)(n+3)}\right) , \tag{14}$$

$$G_2 \sim N\left(3-\frac{6}{n+1}, \frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}\right) . \tag{15}$$

检验假设 H_0：X 为正态总体。记

$$\sigma_1 = \sqrt{\frac{6(n-2)}{(n+1)(n+3)}} , \tag{16}$$

$$\sigma_2 = \sqrt{\frac{24n(n-2)(n-3)}{(n+1)^2(n+3)(n+5)}} \tag{17}$$

$\mu_2 = 3-\dfrac{6}{n+1}, U_1 = G_1/\sigma_1, U_2 = (G_2-\mu_2)/\sigma_2$。当 H_0 为真且 n 充分大时，近似有

$$U_1 \sim N(0,1) \ , \quad U_2 \sim N(0,1) \ . \tag{18}$$

样本偏度 G_1、样本峰度 G_2 分别依概率收敛于总体偏度 ν_1 和总体峰度 ν_2。当 H_0 为真且 n 充分大时，G_1 与 $\nu_1 = 0$ 的偏离不应太大，而 G_2 与 $\nu_2 = 3$ 的偏离不应太大。当 $|U_1|$ 的观察值 $|u_1|$ 或 $|U_2|$ 的观察值 $|u_2|$ 过大时，就拒绝 H_0。取显著性水平为 α，H_0 拒绝域为

$$|u_1| \geqslant k_1, \quad \text{or} \quad |u_2| \geqslant k_2 \tag{19}$$

其中 k_1, k_2 由下式得到

$$P_{H_0}\{|U_1| \geqslant k_1\} \quad = \quad \frac{\alpha}{2}\ , \tag{20}$$

$$P_{H_0}\{|U_2| \geqslant k_2\} \quad = \quad \frac{\alpha}{2}\ . \tag{21}$$

$P_{H_0}\{\cdot\}$ 表示当 H_0 为真时事件 $\{\cdot\}$ 的概率，即有 $k_1 = z_{\alpha/4}, k_2 = z_{\alpha/4}$。拒绝域为

$$|u_1| \geqslant z_{\alpha/4}, \quad \text{or} \quad |u_2| \geqslant z_{\alpha/4} \tag{22}$$

当 n 充分大时，

$$\begin{aligned} P\{\text{当}H_0\text{为真时拒绝}H_0\} &= P_{H_0}\{(|U_1| \geqslant z_{\alpha/4})\bigcup(|U_2| \geqslant z_{\alpha/4})\} \\ &\leqslant P_{H_0}\{|U_1| \geqslant z_{\alpha/4}\} + P_{H_0}\{|U_2| \geqslant z_{\alpha/4}\} \\ &= \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha \end{aligned} \tag{23}$$

样本容量需大于 100。

秩和检验

设有两个连续型总体，概率密度函数分别为 $f_1(x), f_2(x)$，均为未知，但已知

$$f_1(x) = f_2(x-a)\ , \quad a\text{为未知常数}\ , \tag{24}$$

$f_1(x)$ 和 $f_2(x)$ 至多只差一平移。检验下述假设

$$H_0 : a = 0, \quad H1 : a < 0\ . \tag{25}$$

$$H_0 : a = 0, \quad H1 : a > 0\ . \tag{26}$$

$$H_0 : a = 0, \quad H1 : a \neq 0\ . \tag{27}$$

若两总体均值 μ_1, μ_2 存在，由于 f_1, f_2 至多只差一平移，则

$$\mu_2 = \mu_1 - a \ .$$

上述假设分布等价于

$$H_0 : \mu_1 = \mu_2, \quad H1 : \mu_1 < \mu_2 \ . \tag{28}$$

$$H_0 : \mu_1 = \mu_2, \quad H1 : \mu_1 > \mu_2 \ . \tag{29}$$

$$H_0 : \mu_1 = \mu_2, \quad H1 : \mu_1 \neq \mu_2 \ . \tag{30}$$

设 X 为一总体，将一容量为 n 的样本观察值按自小到大的次序编号排列成

$$x_{(1)} < x_{(2)} < \cdots < x_{(n)} \ , \tag{31}$$

称 $x_{(i)}$ 的足标 i 为 $x_{(i)}$ 的秩，$i = 1, 2, \cdots, n$。设自 $1, 2$ 两总体分布抽取容量为 n_1, n_2 的样本，且设两样本独立。假定 $n_1 \leqslant n_2$。将 $n_1 + n_2$ 个观察值放在一起，按自小到大的次序排列，求出每个观察值的秩，将属于第 1 个总体的样本观察值的秩相加，其和记为 R_1，称为第 1 样本的秩和。其余的观察值的秩的总和记作 R_2，称为第 2 样本的秩和。R_1, R_2 是离散型的随机变量，且

$$R_1 + R_2 = \frac{1}{2}(n_1 + n_2)(n_1 + n_2 + 1) \ . \tag{32}$$

R_1, R_2 中一个确定后另一个随之确定。当 H_0 为真时，即有 $f_1(x) = f_2(x)$，两独立样本来自同一个总体。因而第 1 个样本中诸元素的秩应该随机地、分散在自然数 $1 \sim n_1 + n_2$ 中取值，不应过分集中取较小或较大的值。考虑到

$$\frac{1}{2}n_1(n_1 + 1) \leqslant R_1 \leqslant \frac{1}{2}n_1(n_1 + 2n_2 + 1) \ ,$$

即知当 H_0 为真时秩和 R_1 不应取太靠近上述不等式两端的值。因而当 R_1 的观察值 r_1 过分大或过分小时，拒绝 H_0。在给定显著性水平 α 下，H_0 的拒绝域为

$$r_1 \leqslant C_U\left(\frac{\alpha}{2}\right) \quad 或 \quad r_1 \geqslant C_L\left(\frac{\alpha}{2}\right) \ ,$$

其中临界点 $C_U\left(\frac{\alpha}{2}\right)$ 是满足 $P_{a=0}\left\{R_1 \leqslant C_U\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2}$ 的最大整数，而 $C_L\left(\frac{\alpha}{2}\right)$ 是满足 $P_{a=0}\left\{R_1 \geqslant C_L\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2}$ 的最小整数。而犯第 I 类错误的概率为

$$P_{a=0}\left\{R_1 \leqslant C_U\left(\frac{\alpha}{2}\right)\right\} + P_{a=0}\left\{R_1 \geqslant C_L\left(\frac{\alpha}{2}\right)\right\} \leqslant \frac{\alpha}{2} + \frac{\alpha}{2} = \alpha\ .$$

假设检验问题的 p 值检验法

之前讨论的假设检验法称为临界值法。

假设检验问题的p 值(probability value) 是由检验统计量的样本观察值得出的原假设可被拒绝的最小显著性水平。任一检验问题的 p 值可以根据检验统计量的样本观察值以及检验统计量在 H_0 下一个特定的参数值 (一般是 H_0 和 H_1 所规定的参数的分界点) 对应的分布求出。

若显著性水平 $\alpha \geqslant p$，则对应的临界值 $z_\alpha \leqslant$ 观察值 z_0，表示观察值落在拒绝域内，拒绝 H_0；若显著性水平 $\alpha < p$，则对应的临界值 $z_\alpha >$ 观察值 z_0，表示观察值不落在拒绝域内，接受 H_0。p 值是原假设 H_0 可被拒绝的最小显著性水平。

任一检验问题的 p 值可以根据检验统计量的样本观察值以及检验统计量在 H_0 下一个特定的参数值 (H_0 和 H_1 所规定的参数的分界点) 对应的分布求出。

对于任意指定的显著性水平 α，

(1) 若p 值 $\leqslant \alpha$，则在显著性水平 α 下拒绝 H_0。

(2) 若p 值 $\geqslant \alpha$，则在显著性水平 α 下接受 H_0。

用来确定 H_0 的拒绝域。利用 p 值来确定检验拒绝域的方法，称为p 值检验法。

若 p 值 $\leqslant 0.01$，则推断拒绝 H_0 的依据很强或检验是高度显著的；若 $0.01 < p$ 值 $\leqslant 0.05$，则推断拒绝 H_0 的依据强或检验是显著的；若 $0.05 < p$ 值 $\leqslant 0.1$，则推断拒绝 H_0 的理由是弱的，检验是不显著的；若 p 值 > 0.1，一般来说没有理由拒绝 H_0。

[2] “参数估计” 问题是由观测值样本 $\boldsymbol{x}$ 推断分布参数 θ，其中随机变量分布 $p(x;\theta)$ 的函数形式已知。若理论上还不能给出被观测随机变量的概率分布的确切形式，或者虽然存在着某个理论分布公式，但不知道它是否适用于具体的实验观测条件，此时被采用的某个分布形式只是一个假设，它是否合理还需要根据观测得到的样本来判断，即假设检验。

显著性检验

统计假设

检验统计量和显著水平

拟合性检验 $H: p(x;\theta) = f(x;\theta)$

根据样本 $\boldsymbol{x}$ 检验随机变量 $\boldsymbol{x}$ 的分布 $p(x;\theta)$ 是否为某种假定的分布形式 $f(x;\theta)$，即

$$\text{检验 } H: p(x;\theta) = f(x;\theta) \ ,$$

这一类的检验问题叫做拟合性检验。适用于检验任意形式的假设分布 $f(x;\theta)$。但它们几乎都是大样本的检验法，所利用的各种检验统计量 λ，只有在大样本的情况下才渐近服从一个与假设分布 $f(x;\theta)$ 的形式无关的极限分布。对于一种特定的假设分布形式，可以找到比一般方法更精确的检验法——一个更适用的检验统计量，但它不能应用于其他分布形式的检验。

(一) χ^2 检验

(5) 计算皮尔逊 (Pearson)χ^2 量：

$$\chi^2 = \sum_{i=1}^{m} \frac{(n_i - E_i)^2}{E_i} \ . \tag{33}$$

当观测值个数 $N \to \infty$ 时，皮尔逊 χ^2 量渐近地服从自由度 $\nu = m - K - 1$ 的 χ^2 分布，即

$$p\left(\chi^2 = \sum_{i=1}^{m} \frac{(n_i - E_i)^2}{E_i}\right) \to \chi^2(\nu = m - K - 1) \ . \tag{34}$$

因此可以用皮尔逊 χ^2 量作检验。

(6) 利用皮尔逊 χ^2 量作为检验统计量，选取显著水平为 α，由 χ^2 分布表查出 $\chi^2_{1-\alpha}(m - K - 1)$，则拒绝域为

$$\omega = (\chi^2_{1-\alpha}(m - K - 1), \infty) \ ,$$

即

皮尔逊 χ^2 量的大小表现了实测频数 n_i 和理论预期值 E_i 之间差异的大小。

参数检验 (简单假设)

如何由观测值样本来判断关于分布参数的两个假设 (原假设 H 和备择假设 H')。参数检验问题中，观测值 x 的分布 $p(x;\theta)$ 的函数形式已知，对于简单假设的情况，参数 θ 只能取 θ_0 或者 θ_1 这两个值，即对于参数 θ 的原假设和备择假设为

$$H : \theta = \theta_0, \ H' : \theta = \theta_1 \ . \tag{35}$$

“检验 $H : \theta = \theta_0$, $H' : \theta = \theta_1$”，就是在已知参数只可能取 θ_0 和 θ_1 这两个值的情况下，根据观测值样本 $\boldsymbol{x} = (x_1, x_2, \cdots, x_N)$ 判断参数 θ 的值是不是 θ_0。

选择一个检验统计量 $\lambda = \lambda(\boldsymbol{x})$，然后决定一个显著水平 α 的拒绝域 ω，若样本的 λ 值落入拒绝域 ω 内，则拒绝原假设 H。但是，即使 H 为真，λ 值也有一定的概率 (即显

著水平 α) 落入拒绝域内：

$$\begin{aligned} P_r(\lambda \in \omega | H) &= \int_{\lambda\in\omega} p(\lambda|H)\mathrm{d}\lambda \\ &= \int_{\lambda\in\omega} p(\lambda|\theta=\theta_0)\mathrm{d}\lambda \\ &= \alpha \end{aligned} \tag{36}$$

H 为真时错误地拒绝假设 H，叫做假设检验的第一类错误。出现第一类错误的概率就是显著水平 α，又叫做检验的损失。

若 λ 值落在拒绝域 ω 外，则接受假设 H(根据样本 $\boldsymbol{x}$ 没有明显的理由拒绝假设 H)。H 非真而 H' 为真时，λ 值也有一定的概率落到拒绝域之外，即落入区域 $W-\omega$ 内，

$$\begin{aligned} P_r(\lambda \in W-\omega | H') &= 1 - P_r(\lambda \in \omega | H') \\ &= 1 - \int_{\lambda\in\omega} p(\lambda|H')\mathrm{d}\lambda \\ &= 1 - \int_{\lambda\in\omega} p(\lambda|\theta=\theta_1)\mathrm{d}\lambda \\ &= \beta \end{aligned} \tag{37}$$

H 为非真而 H' 为真时，错误地接受假设 H，叫做假设检验的第二类错误。出现第二类错误的概率 β 叫做检验的污染。只有在污染 β 相当小时，所谓 “接受原假设 H” 才意味着原假设很可能是正确的。

同时考虑损失和污染这两类错误的大小，提供了衡量检验方法好坏的标准：为了有效地分辨原假设 H 和备择假设 H'，要求检验的损失 α 和污染 β 都比较小。选定检验的损失，即选定显著水平 α 后，可采用各种不同的统计量作为检验统计量。对于某个检验统计量，又可以采用各种不同的区域作为拒绝域，只要在原假设 H 为真时检验统计量在该区域内的概率含量为 α 即可。如果只考虑检验的损失，则存在着很多种不同的检验方法。在一定的显著水平，即一定的检验损失的要求下，应当采用污染比较小的检验

方法。

$$1-\beta = P_r(\lambda \in \omega | H') = \int_{\lambda\in\omega} p(\lambda|\theta=\theta_1)\mathrm{d}\lambda \ . \tag{38}$$

$1-\beta$ 叫做检验的功效。功效即 H' 被拒绝的份额。对于选定的损失 α，应当采用污染 β 较小，即功效 $1-\beta$ 较大的检验方法。

似然比检验

在损失 α 一定的情况下，污染 β 最小 (功效 $1-\beta$ 最大) 的检验方法叫做佳效检验。对于观测值分布 $p(x;\theta)$ 参数的简单假设

$$H:\theta=\theta_0,\ H':\theta=\theta_1 \ . \tag{39}$$

References

[1] 盛骤 谢式千 潘承毅. 概率论与数理统计. 高等教育出版社, 2008.

[2] 李惕培. 实验的数学处理. 实验物理学丛书. 科学出版社, 1980.