哈爾濱工業大學

第37讲 非参数假设检验
——拟合优度检验

前面讨论了正态总体下，参数的假设检验问题.

实际问题中，可能总体服从何种分布并不知道. 需要我们对总体的分布形式提出假设. 利用数据(样本) 对假设进行检验，看能否获得通过.

例1 在一实验中，每隔一定时间观察一次由某种铀所放射的到达计数器上的 α 粒子数X，共观察了100次，数据如下

α粒子数	0	1	2	3	4	5	6	7	8	9	10	11	≥12
观察次数	1	5	16	17	26	11	9	9	2	1	2	1	0

上述实验数据与X服从泊松分布的理论结果是否相符?　检验方法—χ^2拟合优度检验

χ^2拟合优度检验

检验：$H_0: F(x) = F_0(x)$,

其中，$F(x)$为总体X的分布函数，

$F_0(x)$是某一完全已知或类型已知但含若干

未知参数的分布函数.

当总体X是离散型，

H_0：X的分布列为$P(X = x_i) = p_i (i = 1, 2, \cdots)$.

当总体X是连续型，

H_0：X的概率密度为$f(x)$.

χ^2拟合优度检验

χ^2拟合优度检验的原理与步骤:

1. 将总体X的取值范围分成k个互不相交的区间$A_1, A_2, \cdots, A_k$.

2. H_0成立下，当$F_0(x)$含有未知参数时，求未知参数的最大似然估计值;

3. H_0成立下，可以算出总体X的值落入每个A_i的概率p_i(或 $\hat{p}_i$), 称np_i(或 $n\hat{p}_i$)为落入A_i的样本值的理论频数.

χ^2拟合优度检验

[注] 当$F_0(x)$中含有m个未知参数时，用最大似然法估计未知参数，算得p_i的估计$\hat{p}_i$.

4. 数出A_i中含样本值个数n_i, 并计算统计量

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i},$$

可以证明如下结论

χ^2拟合优度检验

定理 在H_0成立下，当n充分大时，

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-np_i)^2}{np_i}\overset{\text{近似}}{\sim}\chi^2(k-1),$$

$$\chi^2=\sum_{i=1}^{k}\frac{(n_i-n\hat{p}_i)^2}{n\hat{p}_i}\overset{\text{近似}}{\sim}\chi^2(k-m-1),$$

其中，k为子区间个数，m为$F_0(x)$中未知参数的个数.

χ^2拟合优度检验

5. H_0 的显著性水平为 α 的检验的拒绝域为

$$\chi^2 \geq \chi^2_\alpha(k-1),\text{(不需估计未知参数)}$$

$$\chi^2 \geq \chi^2_\alpha(k-m-1),\text{(有}m\text{个未知参数需估计)}$$

注意：该检验方法是在 n 充分大时使用的，因而，**一般要求 $n \geq 50$，及 $np_i \geq 5$.** 否则应适当合并区间，使 np_i 满足要求.

下面用拟合优度检验解例1

解 检验：

$$H_0: P(X=i)=\frac{\lambda^i}{i!}e^{-\lambda}(i=0,1,\cdots).$$

先求出 λ 的最大似然估计 $\hat{\lambda}=\bar{x}=4.2.$

若H_0成立，则$P(X=i)$有估计

$$\hat{p}_i=\hat{P}(X=i)=\frac{4.2^i}{i!}e^{-4.2}(i=0,1,\cdots,11).$$

$$\hat{p}_{12}=\hat{P}(X\geq 12)=1-\sum_{i=0}^{11}\hat{p}_i.$$

计算结果如下：

i	n_i	$\hat{p}_i$	$n\hat{p}_i$	$n_i - n\hat{p}_i$	$\frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$
0	1	0.015	1.5		
1	5	0.063	6.3	−1.8	0.415
2	16	0.132	13.2	2.8	0.594
3	17	0.185	18.5	−1.5	0.122
4	26	0.194	19.4	6.6	2.245
5	11	0.163	16.3	−5.3	1.723
6	9	0.114	11.4	−2.4	0.505
7	9	0.069	6.9	2.1	0.693
8	2	0.036	3.6		
9	1	0.017	1.7		
10	2	0.007	0.7	−0.5	0.039
11	1	0.003	0.3		
12	0	0.002	0.2		
		$\sum$			6.282

检验统计量的值为

$$\chi^2 = \sum_{i=0}^{\infty} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i} = 6.282,$$

在 $\alpha = 0.05$下的临界值为

$$\chi_\alpha^2(k-m-1) = \chi_{0.05}^2(8-1-1) = 12.592.$$

6.282<12.592,故接受H_0，即认为试验数据与理论结果相符.

谢　　谢！