§8.6 拟合优度检验

$$H_0 : F = F_0 \leftrightarrow H_1 : F \neq F_0.$$

- $t_1 < \cdots < t_m$. $1 \ll m \ll n$. $m+1$ 个区间$I_1, \cdots, I_{m+1}$.
- 假设在I_i 中, 有$V_i = v_i$ 个数据.
- SLLN, 概率$\approx$ 频率:

$$\frac{v_i}{n} \approx P(X \in I_i) \overset{H_0}{=} F(t_i) - F(t_{i-1}) = p_i.$$

- CLT:

$$\frac{v_i - np_i}{\sqrt{np_i}} \approx \frac{v_i - np_i}{\sqrt{np_i(1-p_i)}} \text{ 近似服从 } N(0,1).$$

- 一个约束条件: $v_1 + \cdots + v_{m+1} = n$.

- V 近似地服从$\chi^2(m)$:

$$V = \sum_{i=1}^{m+1} \left(\frac{V_i - np_i}{\sqrt{np_i}} \right)^2 = \sum_{i=1}^{m+1} \frac{(V_i - np_i)^2}{np_i} = \sum_{i=1}^{m+1} \left(\frac{V_i}{n} - p_i \right)^2 \frac{n}{p_i}.$$

- 否定域: $\mathcal{W} = \{\vec{x} : V > \chi^2_{1-\alpha}(m)\}$.
- 注1: 结论应为接受H_0.
- 注2: $F_0 = \{F_\theta, \theta = (\theta_1, \cdots, \theta_k) \in \Theta\}$. 检验$H_0 : F \in \mathcal{F}_0$.

  解: 先求最大似然估计$\hat{\theta}$, 再检验$H_0 : F = F_{\hat{\theta}}$, 否定域:

  $\mathcal{W} = \{\vec{x} : V > \chi^2_{1-\alpha}(m-k)\}$.

例6.2. 某计算机产生的一组标准正态随机数: $n = 30$. 检验$H_0 : F = N(0,1)$.

- 分割点: $0, \pm 0.8, \pm 1.6$.

v_i	4	3	11	5	4	3
p_i	0.0548	0.1571	0.2881	0.2881	0.1571	0.0548

- $v = 7.43 < 11.07 = \chi^2_{0.95}(5)$, 接受$H_0$.

2. 柯氏(Kolmogorov)检验法

$$H_0 : F = F_0 \leftrightarrow H_1 : F \neq F_0.$$

- SLLN: 经验分布函数$F_n(x) = \frac{1}{n}\sum_{i=1}^n 1_{\{X_i \leqslant x\}} \approx F(x)$.
- CLT: $\sqrt{n}(F_n(x) - F(x))$ 近似服从$N(0, \sigma^2)$.
- 定理6.1. $D_n = \sup_x |F_n(x) - F(x)|$, 则$\sqrt{n}D \xrightarrow{d} \xi$, 其中

$$F_\xi(x) = \sum_{k=-\infty}^{\infty} (-1)^k \exp(-2k^2x^2) 1_{\{x>0\}} = Q(x).$$

- 将数据$x_1, \cdots, x_n$ 排序得到$x_{(1)} < \cdots < x_{(n)}$. 则,

$$D_n = \max_{1\leqslant k\leqslant n} \max\left\{\frac{k}{n} - F_0(x_{(k)}), F_0(x_{(k)}-) - \frac{k-1}{n}\right\}.$$

- 否定域: $\mathcal{W} = \{\vec{x} : D_n(\vec{x}) > c\}$, 其中$P(\xi > \sqrt{n}c) = \alpha$.