

## 修正极大似然参数估计的问题

Cosslett(1981)那个修正的似然函数

$$l_m(\beta, \theta) = \sum_{j=1}^J \sum_{i=1}^{n_j} \ln \left[ \frac{\lambda_j \Pr(Y = j | G = g_{ij})}{\sum_{k=1}^J \lambda_k \Pr(Y = k | G = g_{ij})} \right]$$

其中 $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{J-1})$ ,  $\lambda_J = n_J/n$ . 总体信息G的边缘概率分布和Y的边缘概率分布, 都不知道的情况下估计参数 $\beta$ 和 $\theta$ . 理论上来说都是 $\beta, \theta$ 都是相合的。

我先给定 $\beta, \theta$ 产生总体。根据你文中的方式一样在从总体中采样出case-control样本。我尝试了几种方式, 估计参数都不是很理想

1. 同时估计 $\lambda, \beta, \theta$ , 这种情况下, 给定初值不同, 有不同的解, 结果也不好。
2. 其次, 我看文章极大似然的步骤, 先极大化 $\lambda$ , 再极大化 $\beta, \theta$ , 此时从结果上来看好像是 $\beta$ 是接近真值的, 如果重复1000次, 感觉比前瞻性似然还要差, 但 $\theta$ 就差很远。而且通常极大化的时候给不同初值, 也有不一样的结果。
3. 我把2中的方法, 不断迭代直至 $\beta, \theta$ 稳定, 但同样有上述问题,  $\beta$ 是接近真值的, 而且效果也不好, 但 $\theta$ 就差很远, 而且极大化给的初值也会有影响
4. 最后我在已知总体信息(产生总体样本中获得的), 计算出理论的 $\lambda_j = \frac{n_j}{N_j}$ , 此时 $\beta, \theta$ 都是相合的, 无偏。但是好像不符合假设条件(Y的边缘分布)。