初值选取和其它参数估计问题 Cosslett(1981)那个修正的似然函数

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

其中$\lambda=(\lambda_1,\lambda_2,\ldots,\lambda_{J-1}),\lambda_J=n_J/n.$

产生数据是已知参数根据proportional odds model，产生总体样本N=100000，比如参数真值为$\theta^0=(3.48,4.6),\beta^0=log(1.4)$。

$$Pr(Y=1)=0.97,Pr(Y=2)=0.02,Pr(Y=1)=0.01$$

然后在从这个总体抽出case-control样本n=1000个,其中每一类为$n_1=500,n_2+n_3=500$。当利用case-control样本估计参数时，总体的信息（Y和G的边缘分布不知道），然后我看Cosslett的文章好像是分步极大化参数，而且理论上λ的真知为$\lambda_j=\frac{n_j/n}{Pr(Y=j)}$,其中$Pr(Y=j)$为总体的概率。

此时我尝试过极大化所有参数，只有β的估计值是相对稳定的，其他比如λ,θ的参数每次都差了不少。给定初值$c(\theta_1,\theta_2,\theta_3,\beta,\lambda_1,\lambda_2,\lambda_3)$(顺序和你的函数有所不同),我用了两个极大化的参数。另$\lambda_J^0=\frac{n_j}{n}$,则根据产生总体的信息(估计参数时并不知道)，此时估计的真值应为$\lambda_j^0=Pr(Y=J)*\frac{n_j/n}{Pr(Y=j)}$

极大化时约束条件$\lambda_j>0,\theta_1<\theta_2<\theta_3$,我给不同初始值时，有时效果好，中位数正好在真值上，有时就差很远。

	constrOptim					maxLix				
初值	β	θ_1	θ_2	λ_1	λ_2	β	θ_1	θ_2	λ_1	λ_2
aa	186	49	14							
Aa	647	293	150							
AA	420	487	260							
Total	1253	829	424							

图b 回顾型数据