多响应近似线性回归模型的 Minmax 稳健设计

刘　欣, 岳荣先

(上海师范大学 数理信息学院, 上海 200234)

摘　要: 研究多响应变量近似线性回归模型的 Minmax 稳健设计问题. 以响应变量均方误差阵的迹 $tr(MSEM)$ 为准则, 对回归方程系数采用广义最小二乘估计, 进而利用最小二乘法得到最优逼近函数. 从而将 Huber (1975) 的方法和 Wiens (1990) 的结果推广到多响应变量场合. 最后考察了一个特例, 即当回归变量间可能存在交互作用时的双响应二元曲面线性回归模型, 得到了与 Wiens (1990) 较一致的结果.

关键词: 多响应模型; 稳健设计; Minmax 设计; 广义最小二乘估计

中图分类号: O212.6　　**文献标识码:** A　　**文章编号:** 1000-5137(2006)01-0036-05

1 前　言

近似线性模型的稳健设计问题最早由 Box Draper (1959) 提出并加以研究, 主要讨论了有限维偏差函数空间 H 上的稳健试验设计问题. 通过分析比较由偏差和波动方差引起的误差的相对重要性, 他们阐明了过于依赖假定的回归函数形式可能导致的问题, 并指出如果一个设计只使由波动方差引起的误差达到最小, 那么即使真实模型与假定模型之间有很小的偏离, 也会使这个设计在波动方差最小准则下所得到的优良性丧失. 在上述研究中, Box Draper 考虑的真实模型是一个比假定的回归函数阶数更高的多项式. Huber (1975) 和 Marcus Sacks (1976) 阐明并解决了在更宽松的限制条件下的 minimax 设计问题, 即: 当真实响应函数在单变量线性函数的邻域内变化时, 寻找设计使最大均方误差达到最小.

但是从稳健性角度来看, Marcus Sacks 选用的领域显然过于狭窄, 以至于他们得到的最优设计只包含两个试验点, 而使得回归函数的线性假定无法进行检验. Wiens (1990) 将 Huber (1975) 的方法推广到多元回归模型并考虑了两种特殊邻域: p 维平面和自变量含交互作用的二元曲面.
本文作者在 Wiens (1900) 工作的基础上, 进一步将 Huber (1975) 的方法推广到多响应变量回归模型.

首先假定真实模型为:

$$y_i = f_i(x) + \varepsilon_i, \quad i = 1, 2, \cdots, r,$$

其中 $(\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_r)^T$ 是均值为 0, 协方差阵为 $\sum$ 的误差向量. 或者用向量形式表示为:

$$y = f(x) + e, \qquad e \sim (0, \Sigma).$$

其中

收稿日期: 2005-11-08

基金项目: 国家自然科学基金(10271078)资助.

作者简介: 刘欣 (1981 -), 男, 上海师范大学数理信息学院硕士研究生; 岳荣先(1958 -), 男, 上海师范大学数理信息学院教授.

$$y=(y_1,y_2,\cdots,y_r)^T,\ f(x)=(f_1(x),f_2(x),\cdots,f_r(x))^T,\ e=(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_r)^T.$$

其次假定试验者在某个经过编码变换后具有单位体积的设计区域 R 内的 n 个设计点 x_i（可以有重复）上做试验，并在如下的向量函数类中寻找 f 的估计量：

$$\mathscr{G}=\{g(x)=\Lambda g(x)\mid g(x)=(g_1,\cdots,g_p)^T,\ g_1,\cdots,g_p\text{为已知的线性独立函数},$$
$$\Lambda=[\alpha_{ij}]_{r\times p},\ \alpha_{ij}\in R,i=1,\cdots,r,\ j=1,\cdots,p\}.$$

而实际情况是只知道 f 属于 G 的某个确定的 L_2 邻域，即对给定的 $\eta\geqslant 0$

$$\mathscr{F}=\{f(x)\mid \inf_{g\in G}\|f-g\|\leqslant\eta\},\tag{1.1}$$

此处，$\|f\|=(\int f(x)^Tf(x)\mathrm{d}x)^{1/2}$ 是 L_2 范数.

那么设计的稳健性要求所选的设计应尽量避免由于采用 $\mathscr{F}$ 到 $\mathscr{G}$ 的近似线性估计所造成的偏离.

这描述了一种实际当中经常遇到的情况，即当我们对一个数据集进行拟合时，虽然我们知道真实的响应往往不是自变量的线性函数，但是我们还是希望能够用线性函数进行近似的拟合，因为通过定义一个非线性函数来拟合给定的数据通常是十分困难的.

在下面的讨论中，我们假定 $\hat{f}$ 是 f 的广义最小二乘估计，并且设计阵是满秩的. 记 $\xi(x)$ 为设计测度. 取损失函数为均方误差阵的迹，即：

$$Q(f,\xi)=tr\int E[f(x)-\hat{f}(x)][f(x)-\hat{f}(x)]^T\mathrm{d}x.\tag{1.2}$$

所谓 Minimax 设计就是要选择一个设计测度 ξ_0，使

$$\sup_f Q(f,\xi_0)=\inf_\xi\sup_f Q(f,\xi).\tag{1.3}$$

本文作者通过寻找（1.3）的鞍点解，即对任意的 $f\in\mathscr{F}$ 和任意的设计测度 ξ 满足：

$$Q(f,\xi_0)\leqslant Q(f_0,\xi_0)\leqslant Q(f_0,\xi).\tag{1.4}$$

的 (f_0,ξ_0) 来得到 Minimax 设计 ξ_0. 为此，在第二节中，先对固定的设计测度 ξ，构造 f 的一个最小二乘最优逼近函数 $f_0\in\mathscr{F}$. 使其满足条件 $Q(f_0,\xi)=\sup_f Q(f,\xi)$. 其构造方法是 Huber（1975）和 Wiens（1990）方法的推广.

另外，虽然对任何特定的类 $\mathscr{G}$，Minimax 设计都可以得到，但是因为 f_0 依赖于 $\mathscr{G}$ 中元素的形式，所以对不同的 $\mathscr{G}$ 设计测度 ξ 不具有统一的形式. 因此，在第三节中，以一种自变量间存在相互作用的双响应二元回归曲面模型为例，给出满足 $Q(f_0,\xi_0)=\inf_\xi Q(f_0,\xi)$ 的设计 ξ_0.

2　确定最小二乘最优逼近函数

根据（1.1），（1.2）. 如果 g_f 是在 L_2 范数意义下，$\mathscr{G}$ 中与 f 距离最近的向量函数，那么 $f-g_f$ 的分量与 $g_f-\hat{f}$ 分量之间正交. 这使得 $Q(f,\xi)$ 可以分解成误差项，偏差项，和方差项，即：

$$\begin{aligned}Q(f,\xi)&=tr\Big(\int[f(x)-g_f(x)][f(x)-g_f(x)]^T\mathrm{d}x+\int[g_f(x)-E\hat{f}(x)][g_f(x)-E\hat{f}(x)]^T\mathrm{d}x+\int var[\hat{f}(x)]\mathrm{d}x\Big)\\&=\int[f(x)-g_f(x)]^T[f(x)-g_f(x)]\mathrm{d}x+\int[g_f(x)-E\hat{f}(x)]^T[g_f(x)-E\hat{f}(x)]\mathrm{d}x+tr\int var[\hat{f}(x)]\mathrm{d}x\\&\hat{=}Q_f+B(f,\xi)+trV(\xi).\end{aligned}\tag{2.1}$$

记 $z(x)=(g_1(x),\cdots,g_p(x))^T$，$Z=diag(z(x),\cdots,z(x))_{rp\times r}=I_r\otimes z(x)$，$a=\int Zf(x)\mathrm{d}x$，$A=\int ZZ^T\mathrm{d}x$，$\alpha^0=A^{-1}a$. 由最小二乘法得到：

$$g_f(x)=Z^T\alpha^0.\tag{2.2}$$

类似地，记

$$\hat{b}=\int Z\Sigma^{-1}Yd\xi(x),\quad B=\int Z\Sigma^{-1}Z^Td\xi(x),\quad \hat{\alpha}=B^{-1}\hat{b},$$

得到$f(x)$ 的广义最小二乘估计

$$\hat{f}(x) = Z^T\hat{\alpha}. \tag{2.3}$$

由上,记

$$b(f) \hat{=} b \hat{=} E\hat{b} = \int Z\Sigma^{-1}f(x)d\xi(x), \quad \alpha^1 = E\hat{\alpha} = B^{-1}b, \tag{2.4}$$

得到

$$E\hat{f}(x) = Z^T\alpha^1. \tag{2.5}$$

利用 (2.2) ~ (2.5)式 ,再结合 (2.1)式,得到

$$B(f,\xi) = (\alpha^0 - \alpha^1)^T A(\alpha^0 - \alpha^1), \quad trV(\xi) = \frac{1}{n}trAB^{-1}.$$

注意在我们的假设下,A 和 B 都是正定的.

此外,对给定的ξ如果令 $h \hat{=} f - g_f$,那么 $g_h(x) \equiv 0, Q_f = Q_h = \|h\|^2$,再由 (2.2) , $B(f,\xi) = B(h,\xi)$. 而显然 $trV(\xi)$ 不变 . 因此, $Q(f,\xi) = Q(h,\xi)$. 从而,不失一般性,假设 $g_f = 0$. 并记 $H = BA^{-1}B$. 那么问题归结为: 在约束条件:

$$(i)\int Zf(x)\mathrm{d}x = 0, \ (ii)\int f^T f(x)\mathrm{d}x = \eta^2. \tag{2.6}$$

下,求f_0 使 $B(f,\xi) = b^T H^{-1} b$ 达到最大值.

之所以要求满足条件 (*ii*) ,是因为如果f满足 (*i*) ,但 $\|f\| = c\eta < \eta$,那么 $c^{-1}f$ 满足 (*i*) ,并且 $B(c^{-1}f,\xi) = c^{-2}B(f,\xi) > B(f,\xi)$.

下面根据[5]中引理1 将ξ限制在绝对连续测度范围内,并设 $\xi' = m$. 但需要注意的是,实际当中, ξ 总是用离散测度近似代替. 记 $\mathscr{F}_1 = \{f: \|f\| \le \eta\}$,对任意的$f_0, f_1 \in \mathscr{F}_1$ 和 $\lambda \in [0,1]$,令$f_\lambda = (1-\lambda)f_0 + \lambda f_1$. 取 $d \in R^{n'}, d_0 \in R$ 为 Lagrange 乘子, 并令:

$$\phi(\lambda) = B(f_\lambda,\xi) + d^T\int Zf_\lambda \mathrm{d}x + d_0\int f_\lambda^T f_\lambda(x)\mathrm{d}x.$$

从而,对任意的f_1 ,使 $B(f_\lambda,\xi)$ 达到最大的函数f_0 应满足

$$0 \ge \phi'(0) = \int[2b^T H^{-1} Z\Sigma^{-1} m(x) + d^T Z + 2d_0 f_0^T](f_1 - f_0)(x)\mathrm{d}x.$$

此处, $b = b(f_0)$. 上述不等式说明应有

$$f_0(x) = m(x)\Sigma^{-1}Z^T H^{-1}\beta + Z^T c,$$

其中$\beta \in R^{n'}$ [正比于 $b(f_0)$] ,c 为常数向量. 选取 c,使f_0 满足(2.6)(i)得到

$$f_0(x) = [m(x)\Sigma^{-1}Z^T H^{-1} - Z^T B^{-1}]\beta. \tag{2.7}$$

为了证明上述使偏差达到最大的f_0 确实存在,并具有形式(2.7) ,首先定义函数

$$h(x,\beta) = s[m(x)\Sigma^{-1}Z^T H^{-1} - Z^T B^{-1}]\beta,$$

其中 $s > 0$,且使

$$\|h(\cdot,\beta)\| = \eta, \tag{2.8}$$

则对任意的β , $h(x,\beta)$ 都满足 (2.6)(i),(ii) .

下面先证明对于任何给定的满足(2.6)(i),(ii)的函数f, 及相应的 $h(x,b(f))$,都有不等式

$$B(f,\xi) \le B(h(\cdot,b(f)),\xi), \tag{2.9}$$

成立. 从而我们只需要在具有形式(2.7)的函数中寻找f_0 即可. 此外,(2.9)中的等号成立当且仅当

$$f(x) = h(x,b(f)) \quad \text{a.e.} \tag{2.10}$$

记 $K \hat{=} \int Z\Sigma^{-1}Z^T m^2(x)\mathrm{d}x$, 由(2,8)得

$$s^2 = \frac{\eta^2}{b^T(f)[H^{-1}KH^{-1} - H^{-1}]b(f)}. \tag{2.11}$$

另外，用 $h(\cdot,b(f))$ 替换(2.4)的 f，则有

$$b(h(\cdot,b(f))) = s[KH^{-1} - I]b(f), \tag{2.12}$$

进而得到

$$B(h(\cdot,b(f))) = s^2 b^T(f)[H^{-1}K - I]H^{-1}[H^{-1}K - I]b(f). \tag{2.13}$$

又因为 f 满足(2.6)(i) 和 (ii)，所以

$$s^2[b^T(f)H^{-1}b(f)]^2 = (\int f^T(x)h(x,b(f))\mathrm{d}x)^2 \leqslant \int f^T(x)f(x)\mathrm{d}x\int h^T(x,b(f))h(x,b(f))\mathrm{d}x = \eta^4$$

令 $H^{-1/2}$ 表示 H^{-1} 的对称平方根阵，并取 $L = H^{-1/2}b(f)$，$J = H^{-1/2}KH^{-1/2} - I$. 进一步将 J 进行分解，使 $J = Q\Lambda Q^T$，其中 Q 是正交阵，Λ 是对角阵. 再定义 $u = Q^TL/(L^TL)^{1/2}$. 注意有 $u^Tu = 1$. 并记 $\bar{\lambda} = \Sigma_{i=1}^{rN}\lambda_i u_i^2$，其中 $\lambda_i = \Lambda_{ii}$ 为 Λ 的第 i 个对角元. 在上述记号下，我们有

$$B(h(\cdot,b(f)),\xi) = s^2L^TJ^2L,\quad B(f,\xi) = L^TL. \tag{2.15}$$

再由 (2.11) 和 (2.14) 分别得到，

$$s^2 = \eta^2/L^TJL \quad 和 \quad \eta^2 \geqslant sL^TL, \tag{2.16}$$

所以

$$\eta^2 \geqslant \frac{(L^TL)^2}{L^TJL} = \frac{L^TL}{u^T\Lambda u}. \tag{2.17}$$

进而由 (2.15) ~ (2.17) 得到

$$\frac{B(h(\cdot,b(f)),\xi)}{B(f,\xi)} - 1 = \frac{\eta^2}{L}{}^T L \cdot \frac{u^T\Lambda^2}{uu^T\Lambda u} - 1 \geqslant \frac{u^T\Lambda^2}{uu^T\Lambda u^2} - 1 = \frac{\Sigma_{i=1}^{rN}(\lambda_i - \bar{\lambda})^2u_i^2}{\bar{\lambda}^2} \geqslant 0. \tag{2.18}$$

这样就证明了(2.9). 而要使(2.18)中等号成立，则要求(2.14)中等号成立，也就是要使(2.10) 成立. 反之，若(2.10)满足，则显然(2.18)等号成立. 所以(2.9)中等号成立等价于(2.10)满足.

由于确定具有形式(2.7)的 f_0 要比研究(2.10)的解更简单，所以考虑在约束条件

$$\int f_0^T(x)f_0(x)\mathrm{d}x = \beta^T(H^{-1}KH^{-1} - H^{-1})\beta = \eta^2 \tag{2.19}$$

下，选取 β，使 $B(f_0,\xi) = \beta^T(KH^{-1} - I)^TH^{-1}(KH^{-1} - I)\beta$ 达到最大.

若记 $\delta = K^{-1}\beta$，$G = (KH^{-1} - I)KH^{-1}K$，$F = G(H^{-1}K - I)$，则上述问题等价为，在约束条件

$$\delta^TG\delta = \eta^2 \tag{2.20}$$

下，求 δ 使

$$B(f_0,\xi) = \delta^TF\delta \tag{2.21}$$

达到最大.

这就转化为一个标准的特征值问题. 可以先求出满足 $|F - \mu G| = 0$ 的最大特征根 μ_ξ，然后求出与之相对应的，同时满足(2.20)的特征向量，就是通过下式：

$$(F - \mu_\xi G)\delta = 0$$

和(2.20)式求解 δ . 或者，由于 $\nu_\xi = 1 + \mu_\xi$ 是满足：

$$|K - \nu H| = 0 \tag{2.23}$$

的最大特征根，从而等价地，可通过下式：

$$(K - \nu_\xi H)\delta = 0 \tag{2.24}$$

和(2.20)式求解 δ . 再由(2.7)，得：

$$f_0(x) = \nu_\xi[m(x)\Sigma^{-1}Z^T - Z^TA^{-1}B]\delta,$$

至此证明了 f_0 是存在的，并具有形式(2.7). 同时得到最大偏差 $B(f_0,\xi) = \mu_\xi\eta^2$，和

$$Q(f_0,\xi) = \nu_\xi\eta^2 + \frac{1}{n}trAB^{-1}. \tag{2.25}$$

3 例 子

本节针对第一节中提到的一种特殊响应曲面,利用变分方法,求出使(2.25)达到最小的,进而满足奇异性(1.4)的设计 ξ_0 所具有的密度函数 $m_0(x)$.

取 $\mathscr{G}$为 $R=[-\frac{1}{2},\frac{1}{2}]\times[-\frac{1}{2},\frac{1}{2}]$ 上的具有如下形式:

$$g(x_1,x_2)=\alpha_1+\alpha_2x_1+\alpha_3x_2+\alpha_4x_1x_2$$

的函数构成的类.并设只有两个响应,即 $r=2$,且协方差矩阵为 $\Sigma=\begin{pmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{pmatrix}$. 则有 $x=(x_1,x_2)^T$, $z=(1,x_1,x_2,x_1x_2)^T$,$p=4$, $Z=\mathrm{diag}(z,z)_{8\times2}=I_2\otimes z$. 限定 $m(x)$ 关于每个变量都是对称的且变量之间可交换次序. 经计算得到:

$$A=I_2\otimes\mathrm{diag}(1,\frac{1}{12},\frac{1}{12},\frac{1}{144}),\quad B=\Sigma^{-1}\otimes diag(1,\gamma,\gamma,\gamma_{12}),$$

其中:

$$\gamma=\int x_1^2m(x)\mathrm{d}x,\quad \gamma_{12}=\int x_1^2x_2^2m(x)\mathrm{d}x;$$

$$K=\Sigma^{-1}\otimes\mathrm{diag}(\int m^2(x)\mathrm{d}x,\int x_1^2m^2(x)\mathrm{d}x,\int x_1^2m^2(x)\mathrm{d}x,\int x_1^2x_2^2m^2(x)\mathrm{d}x),$$

$$\nu_\xi=\max\left[\int m^2(x)\mathrm{d}x,\frac{\int x_1^2m^2(x)\mathrm{d}x}{12\gamma^2},\frac{\int x_1^2x_2^2m^2(x)\mathrm{d}x}{144\gamma_{12}^2}\right].$$

在下面(3.5)式范围之内,有

$$\nu_\xi=\int m^2(x)\mathrm{d}x,\tag{3.1}$$

由此

$$Q(f_0,\xi)=\eta^2\int m^2(x)\mathrm{d}x+\frac{\sigma_1^2+\sigma_2^2}{n}(1+\frac{1}{6\gamma}+\frac{1}{144\gamma_{12}}).\tag{3.2}$$

为了使 $Q(f_0,\xi)$ 达到最小值,先对固定的 γ,γ_{12} 使其取最小值,然后再关于 γ,γ_{12} 使其达到最小. 第一步:通过引入 Lagrange 乘子 λ,μ,δ,逐点最小化被积函数,使积分

$$\int[m^2(x_1,x_2)-2\lambda\, m(x_1,x_2)-2\mu\,(x_1^2+x_2^2)m(x_1,x_2)-2\delta\, x_1^2x_2^2m(x_1,x_2)]\mathrm{d}x_1\mathrm{d}x_2$$

达到最小. 由此得到

$$m_0(x_1,x_2)=(\lambda+\mu(x_1^2+x_2^2)+\delta\, x_1^2x_2^2)^+,\tag{3.3}$$

其中 λ,μ,δ 由下列等式确定

$$\int m_0(x)\mathrm{d}x=1,\ \int x_1^2m_0(x)\mathrm{d}x=\gamma,\ \int x_1^2x_2^2m_0(x)\mathrm{d}x=\gamma_{12}.\tag{3.4}$$

下面对应于:

$$\frac{1}{12}\leqslant\gamma\leqslant\frac{7}{60}.\tag{3.5}$$

由 (3.4)具体解出 λ,μ,δ. 注意此时它们都是非负的:

$$\lambda=\frac{9}{16}(400\gamma_{12}-120\gamma+9),$$

$$\mu=\frac{45}{4}(-240\gamma_{12}+56\gamma-3),$$

$$\delta=225(144\gamma_{12}-24\gamma+1).$$

代入(3.3)即得到 $m_0(x)$ 关于 $\gamma,\gamma_{12},(\sigma_1^2+\sigma_2^2)/n\eta^2$ 的表达式.

第二步:首先可以得到 γ_{12} 关于 γ 的表达式,进而得到 $(\sigma_1^2+\sigma_2^2)/n\eta^2$ 关于 γ 的表达式.具体地,由下式解得 γ_{12}:

$$34560\gamma_{12}^3+240(1-24\gamma)\gamma_{12}^2+\gamma^2(3-56\gamma)=0,$$

并取满足下式的根:

$$\frac{1}{6}\gamma-\frac{1}{144}\gamma_{12}\frac{7}{30}\gamma-\frac{1}{80},$$

以保证 $\lambda,\mu,\delta\geqslant 0$. 然后代入下式:

$$(\sigma_1^2+\sigma_2^2)/n\eta^2=270\gamma^2(56\gamma-240\gamma_{12}-3),$$

解得 $(\sigma_1^2+\sigma_2^2)/n\eta^2$,再代入(3.3)即可得到我们所要求的设计密度函数 $m_0(x)$,同时得到:

$$f_0(x)=\eta\frac{m_0(x)-1}{(\lambda+2\mu\gamma+\delta\gamma_{12})^{1/2}}. \tag{3.6}$$

特别地 $\gamma=\frac{1}{12}$,对应于均匀设计.相关的具体数值可参见 D. P. Wiens. (1990) 表 1.

参考文献:

[1] BOX G E P, DRAPER N R. A basis for the selection of a response surface dasign[J]. J Amer Statist Assoc,1959,54: 622 - 654.

[2] HUBER P. Robustness and designs, in A Survey of statistical Design and Linear Models(Srivastava,J N, Ed)[M]. North Holland,Amsterdam,1975. 287 - 303.

[3] MARCUS M B, SACKS J. Robust designs for regression problems, in Statistical Decision Theory and Related Toipcs II(S S Gupta and D S Moore,Eds.)[M]. New York: Academic, 1976. 245 - 268.

[4] WIENS D P. Robust Minimax Designs for Multiple Linear Regression[J]. Linear Algebra and its applications,1990,127: 327 - 340.

[5] WIENS D P. Minimax designs for approximately linear regression[J]. Journal of Statistical Planning and Inference,1992, 31: 353 - 371.

Robust minimax designs for multiresponse linear regression

LIU Xin, YUE Rong-xian

(Mathmatics and Sciences College, Shanghai Normal University, Shanghai 200234, China)

Abstract: We establish an extension, in the case of multiresponse regression, of a result on minimax multiple regression designs due to D. P. Wiens(1990). A design is found which is minimax with respect to the trace of the integrated mean squared error matrix as the true response function varies over an L_2- neighbourhood of a bivariate surface with possible interactions between the regressors.

Key words: robust minimax designs; multireponse; generalized least estimate