均方误差意义下的无偏估计与有偏估计

景英川[1],项明寅[2]
(1.太原理工大学 数学系,山西 太原 030024;2.黄山学院 数学系,安徽 黄山 245041)

摘要:讨论均方误差意义下的无偏估计与有偏估计,得出线性模型中一些有偏估计在 MSE 准则下具备比无偏估计更好的优良性。

关键词:均方误差;无偏的;参数估计
中图分类号:O212.1 **文献标识码**:A **文章编号**:1672-447X(2009)05-0013-03

1 引 言

估计量的无偏性是参数估计理论中一个重要的概念,也是衡量估计量优劣的标准之一。但是无偏估计不一定存在,且一般不唯一,也不一定比有偏估计更优。因此,我们给出均方误差意义下的无偏估计与有偏估计。从直观上看,一个好的估计应该离真值比较近,在真值的周围波动,同时一个好的估计应该有较小的均方误差。故此,均方误差是评价一个估计优劣的较有力的标准。[1]在本文中,罗列了几种模型之下无偏估计跟有偏估计在均方误差意义下的比较。

2 无偏估计与有偏估计的均方误差比较

考虑 Gauss-Markov 模型

$$\begin{cases} Y = X\beta + \varepsilon \\ E(\varepsilon)=0 \quad \mathrm{cov}(\varepsilon)=\sigma^2 I_n \end{cases} \tag{1}$$

其中$Y=Y_{n\times1}$是因变量的观察值向量, $X = X_{n\times p}$ 为列满秩的设计矩阵, $\beta=\beta_{p\times1}$ 是未知的参数向量, $\varepsilon=\varepsilon_{n\times1}$ 是随机误差向量。最小二乘估计是线性无偏估计类中方差最小的,有偏估计可以在均方意义下优于最小二乘估计,使得这个估计在有偏估计类中是关于均方误差最小。

2.1 岭估计

对于线性模型 (1), 回归系数β的岭估计定义为:

$$\hat{\beta}(k)=(X'X+kI)^{-1}X'Y$$

这里$k>0$是可选择参数,称为岭参数。岭估计是有偏估计,但是其均方误差很小,在复共线性存在的情况下,岭估计优于最小二乘估计。为证明其优良性质,引入模型(1)的典则形式:

设$\lambda_1 \ge \cdots \ge \lambda_p$为$X'X$的特征根, $\varphi_1,\cdots,\varphi_p$为对应的标准正交化特征向量。记 $\Phi=(\varphi_1,\cdots,\varphi_p)$为 $p\times p$ 正交阵,再记 $\Lambda = diag(\lambda_1,\cdots,\lambda_p)$,于是 $X'X=\Phi\Lambda\Phi'$。则线性回归模型可改写为:

$$\begin{cases} Y=\alpha_0 1_p + Z\alpha + \varepsilon \\ E(\varepsilon)=0 \quad \mathrm{cov}(\varepsilon)=\sigma^2 I_n \end{cases} \tag{2}$$

其中常数项记为α_0,$Z=X\Phi$,$\alpha=\Phi'\beta$。

典则回归系数的最小二乘估计 $\hat{\alpha}=\Lambda^{-1}Z'Y$,岭估计为 $\hat{\alpha}(k)=(\Lambda+kI)^{-1}Z'Y$,另外还有估计 $\hat{\alpha}_k=(\Lambda+kI)(Z'Y+\hat{\alpha})$。可知$\hat{\alpha}_k$是$\hat{\alpha}$的一个线性变换,且当$k>1$时,$\hat{\alpha}_k$是$\hat{\alpha}$向原点的压缩。

容易证明以下结论成立:

收稿日期:2009-06-20
作者简介:景英川(1970-),山西临汾人,太原理工大学数学系副教授,硕士,研究方向为概率统计。

$MSE(\hat{\alpha})=MSE(\hat{\beta})$, $MSE(\hat{\alpha}(k))=MSE(\hat{\beta}(k))$ 。

定理 2.1.1　存在 $k>0$,使得

$MSE(\hat{\beta}(k))<MSE(\hat{\beta})$ 。

证明:只需证明 $MSE(\hat{\alpha}(k))<MSE(\hat{\alpha})$

因为设计阵 Z 中心化,有 $1'Z=0$。所以

$$E\hat{\alpha}(k)=(\Lambda+kI)^{-1}Z'(\alpha_0 1+Z\alpha)=(\Lambda+kI)^{-1}Z'Z\alpha$$
$$=(\Lambda+kI)^{-1}\Lambda\alpha$$
$$Cov\hat{\alpha}(k)=\sigma^2(\Lambda+kI)^{-1}Z'Z(\Lambda+kI)^{-1}$$
$$=\sigma^2(\Lambda+kI)^{-1}\Lambda(\Lambda+kI)^{-1}$$

于是

$$MSE(\hat{\alpha}(k))=trCov\hat{\alpha}(k)+\|E\hat{\alpha}(k)-\alpha\|^2$$
$$=\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^2}+k^2\sum_{i=1}^{p}\frac{\sigma_i^2}{(\lambda_i+k)^2}$$
$$=f_1(k)+f_2(k)=f(k)。$$

对 k 求导,得

$$f_1'(k)=-2\sigma^2\sum_{i=1}^{p}\frac{\lambda_i}{(\lambda_i+k)^3},\ f_2'(k)=-2\sigma^2\sum_{i=1}^{p}\frac{\lambda_i\alpha_i^2}{(\lambda_i+k)^3}。$$

因为

$$f_1'(0)=-2\sigma^2\sum_{i=1}^{p}\frac{1}{\lambda_i^2}<0,\qquad f_2'(0)=0。$$

于是 $f'(0)<0$,但是 $f'(k)$ 在 $k\geq 0$ 时候都连续,故 $f'(k)$ 在 $k>0$ 且充分小时 $f'(k)<0$。亦即 $f(k)=MSE(\hat{\alpha}(k))$ 在 $k(>0)$ 充分小时是 k 的单调减函数,因而存 $k^*>0$ 在 $k\in(0,k^*)$ 时有 $f(k)<f(0)=MSE(\hat{\alpha})$,这就证明了 $MSE(\hat{\alpha}(k))<MSE(\hat{\alpha})$。

2.2　在实际情况中,线性模型的误差并不总是等方差且不相关的

接下来我们把以上估计推广到模型

$$\begin{cases}Y=X\beta+\varepsilon\\ E(\varepsilon)=0\quad \operatorname{cov}(\varepsilon)=\sigma^2V,\ V>0\end{cases}\tag{3}$$

给出 β 的广义最小二乘估计 $\beta^*=(X'V^{-1}X)^{-1}X'V^{-1}Y$,可以计算 β 的广义岭型估计 $\beta^*(k)=(X'V^{-1}X+kI)^{-1}X'V^{-1}Y$。另外还有估计 $\beta^*_k=(X'V^{-1}X+kI)^{-1}(X'V^{-1}X+\beta^*)$。只要 $k\neq 1,\beta^*(k)$ 以及 β^*_k 皆为有偏估计,为便于讨论这种估计的性质,对模型(3)作适当的变换:

$$\tilde{Y}=V^{-\frac{1}{2}}Y;\tilde{X}=V^{-\frac{1}{2}}X;\tilde{\varepsilon}=V^{-\frac{1}{2}}。$$

得到

$$\begin{cases}\tilde{Y}=\tilde{X}\beta+\tilde{\varepsilon}\\ E(\tilde{\varepsilon})=0\quad \operatorname{cov}(\tilde{\varepsilon})=\sigma^2I\end{cases}\tag{4}$$

模型(4)的典则形式:设 $\lambda_1\geq\cdots\geq\lambda_p$ 为 $\tilde{X}'\tilde{X}=X'V^{-1}X$ 的特征根,$\varphi_1,\cdots,\varphi_p$ 为对应的标准正交化特征向量。记 $\Phi=(\varphi_1,\cdots,\varphi_p)$ 为 $p\times p$ 正交阵,再记 $\Lambda=diag(\lambda_1,\cdots,\lambda_p)$。则线性回归模型可改写为:

$$\begin{cases}\tilde{Y}=\alpha_0 1_p+Z\alpha+\tilde{\varepsilon}\\ E(\tilde{\varepsilon})=0\quad \operatorname{cov}(\tilde{\varepsilon})=\sigma^2I_n\end{cases}\tag{5}$$

其中常数项记为 $\alpha_0,Z=\tilde{X}\Phi,\alpha=\Phi'\beta$。

典则回归系数的广义最小二乘估计为 $\alpha^*=(\Lambda)^{-1}Z'\tilde{Y}$。岭估计为 $\alpha^*(k)=(\Lambda+kI)^{-1}Z'\tilde{Y}$,另外还有估计 $\alpha^*_k=(\Lambda+kI)^{-1}(Z'\tilde{Y}+\alpha^*)$。可知 α^*_k 是 α^* 的一个线性变换,且当 $k>1$ 时,α^*_k 是 α^* 向原点的压缩。相应的

$$\beta^*=\Phi\alpha^*,\beta^*(k)=\Phi\alpha^*(k),\beta_k^*=\Phi\alpha_k^*。$$

定理 2.2.1　存在时 $k>1$,$MSE(\alpha^*_k)<MSE(\alpha^*)$。

证明:因为

$$\alpha^*_k=(\Lambda+kI)^{-1}(Z'\tilde{Y}+\alpha^*)$$
$$Cov(\alpha^*_k)=\sigma^2(\Lambda+kI)^{-1}(\Lambda+I)\cdot\Lambda^{-1}(\Lambda+I)(\Lambda+kI)^{-1},$$
$$\|E(\alpha_k^*)-\alpha\|=\|(\Lambda+kI)^{-1}(\Lambda+I)\alpha-\alpha\|^2=\|((\Lambda+kI)^{-1}(\Lambda+I)-I)\alpha\|^2$$
$$=\sum_{i=1}^{p}\left(\left(\frac{\lambda_i+1}{\lambda_i+k}-1\right)\alpha_i\right)^2=(1-k)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^2}$$

所以

$$MSE(\alpha_k^*)=tr\operatorname{cov}(\alpha_k^*)+\|E(\alpha_k^*)-\alpha\|=\sigma^2\sum_{i=1}^{p}\frac{(\lambda_i+1)^2}{\lambda_i(\lambda_i+k)^2}$$
$$+(1-k)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^2}\doteq g(k)。$$

可求得

$$g'(k)=-2\sigma^2\sum_{i=1}^{p}\frac{(\lambda_i+1)^2}{(\lambda_i+k)^3}-2(1-k)\cdot\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^2}$$
$$-2(1-k)^2\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_i+k)^3}$$

因为

$$g'(1)=-2\sigma^2\sum_{i=1}^{p}\frac{(\lambda_i+1)^2}{\lambda_i(\lambda_i+k)^3}<0$$ 且 $g(k)$ 在 $k>1$ 时连续,所以存在 $k>1$ 使得 $g(k)<g(1)$,即 $MSE(\alpha^*_k)<MSE(\alpha^*)$。

定理 2.2.2　当 $k>\frac{1}{2}\left\{1+\sigma^2(\alpha'\alpha)^{-1}\sum_{i=1}^{p}\left(2+\frac{1}{\lambda_i}\right)\right\}$ 时,有 $MSE(\alpha^*_k)<MSE(\alpha^*(k))$。

证明:因为

$$\alpha_k^*=\frac{1}{2}(\Lambda+hI)^{-1}(Z^T\tilde{Y}+\alpha^*)=(\Lambda+hI)^{-1}Z^T\tilde{Y}$$
$$+(\Lambda+hI)^{-1}\alpha^*=\alpha^*(h)+(\Lambda+hI)^{-1}\alpha^*,$$

$$D(\alpha^*)=\sigma^2\Lambda^{-1},D(\alpha_{(i)}^*)=\frac{\sigma^2}{\lambda_i},\ E(\alpha_{(i)}^{*\,2})=\frac{\sigma^2}{\lambda_i}+\alpha_i^2,$$

$$MSE(\alpha_k^*)=\sum_{i=1}^{p}E\left(\alpha_{k(i)}^*-\alpha_i\right)^2。$$

又因为

$$E\left(\alpha_{k(i)}^*-\alpha_i\right)^2=E\left[\alpha_{(i)}^*(k)-\alpha_i+\frac{\alpha_{(i)}^*}{\lambda_i+k}\right]^2$$

$$=E(\alpha_{(i)}^*(k)-\alpha_i)^2+2E((\alpha_{(i)}^*(k)-\alpha_i)\cdot\frac{\alpha_{(i)}^*}{\lambda_i+k})+\frac{1}{(\lambda_i+k)^2}E(\alpha_{(i)}^{*2})$$

$$=E(\alpha_{(i)}^*(k)-\alpha_i)^2+\frac{2}{\lambda_i+k}(\frac{\lambda_i}{\lambda_i+k}\cdot\left(\frac{\sigma^2}{\lambda_i}+\alpha_i^2\right)-\alpha_i^2)+\frac{1}{(\lambda_i+k)^2}\left(\frac{\sigma^2}{\lambda_i}+\alpha_i^2\right)$$

$$=E(\alpha_{(i)}^*(k)-\alpha_i)^2+\frac{(2\lambda_i+1)\sigma^2}{\lambda_i(\lambda_i+k)^2}+\frac{1-2k}{(\lambda_i+k)^2}\alpha_i^2。$$

所以，对$i=1,\dots,p$求和，由上式可得

$$MSE(\alpha_k^*)-MSE\left(\alpha^*(k)\right)=\sum_{i=1}^{p}\frac{1-2k}{(\lambda_i+k)^2}\alpha_i^2+\sum_{i=1}^{p}\frac{(2k+1)\sigma^2}{\lambda_i(\lambda_i+k)^2}$$

$$\le(1-2k)\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_p+k)^2}+\sigma^2\sum_{i=1}^{p}\frac{2+\frac{1}{\lambda_i}}{(\lambda_p+k)^2}。$$

因此，要使 $MSE(\alpha^*_k)<MSE(\alpha^*(k))$，只要

$$(1-2k)\sum_{i=1}^{p}\frac{\alpha_i^2}{(\lambda_p+k)^2}+\sigma^2\sum_{i=1}^{p}\frac{2+\frac{1}{\lambda_i}}{(\lambda_p+k)^2}<0。$$

因此当

$$k>\frac{1}{2}\left\{1+\sigma^2(\alpha'\alpha)^{-1}\sum_{i=1}^{p}\left(2+\frac{1}{\lambda_i}\right)\right\}时，$$

$$MSE\left(\alpha^*_k\right)<MSE\left(\alpha^*(k)\right)。$$

2.3 约束岭估计

如果给模型(1)加一个非其次线性等式约束，即模型变为

$$\begin{cases}Y=X\beta+\varepsilon\\ E(\varepsilon)=0\quad \mathrm{cov}(\varepsilon)=\sigma^2 I_n\\ R\beta=r\end{cases}\qquad(6)$$

其中R为$m\times p$非零已知矩阵，r为$m\times1$向量，秩$(R)=m$。

定理 2.3.1　对于线性模型 (6) 设 R为 $m\times p$矩阵，秩$(R)=m$，$M(R')\subset M(X')$，且$R\beta=r$相容，若秩$(X)=P$，则 $b(R)=b+S^{-1}R'(RS^{-1}R')^{-1}(r-Rb)$为$\beta$的约束最小二乘估计。

证明略。

定理 2.3.2　对于线性模型 (6) 设 R为 $m\times p$矩阵，秩$(R)=m$，$M(R')\subset M(X')$，且$R\beta=r$相容，则β的约束型岭估计为：

$$b(k,R)=b(k)+S_k^{-1}R'(RS_k^{-1}R')^{-1}(r-Rb(k))$$

其中 $b(k)=S^{-1}{}_k$ ，$S_k=X'X+kI$

证明略。

定理 2.3.3　当 $0<k\le 2\sigma^2/\beta'U^+\beta$ 时，在均方意义下，约束性岭估计$b(k,R)$优于约束性最小二乘估计$b(R)$，其中$U=I_p-R'(RR')^{-1}R$ ，U^+为U的加逆号。

证明略。

3 结束语

可见，线性模型中一些有偏估计在准则下具备比无偏估计更好的优良性。关于有偏估计，在不同的模型下有不同的形式，用均方误差去评估其好坏，最主要一点是确定k的取值，关于这方面的研究还在继续，文献[2,3]大量运用数学知识推导，给出k的不同取值下均方误差的比较，然后得出较优的估计。

参考文献：

[1]刘冬喜.从约束型最小二乘估计到约束型岭估计的探索[J].湖南工业大学学报,2008,(3):40-42.

[2]赵天玺,李兆勤.回归系数的一种有偏估计[J].重庆工商大学学报(自然科学版),2007,(12):541-614.

[3]林路.回归系数的综合岭估计[J].数理统计与应用概率,1996,(3):179-184.

责任编辑：胡德明

Unbiased Estimation and Biased Estimation in the Sense of the Mean Square Error

Jing Yingchuan[1], Xiang Mingyin[2]

(1. Department of Mathematics, Taiyuan University of Technology, Taiyuan 030024, China;

2. Department of Mathematics, Huangshan University, Huangshan 245041, China)

Abstract: This paper discusses unbiased estimation and biased estimation in the sense of the mean square error. In linear models, some biased estimations are superior to unbiased estimation under MSE criterion.

Key words: mean square error; unbiased; parameter estimation

均方误差意义下的无偏估计与有偏估计

作者：景英川， 项明寅
作者单位：景英川(太原理工大学数学系,山西太原,030024)， 项明寅(黄山学院教学系,安徽黄山,245041)
刊名：黄山学院学报
英文刊名：JOURNAL OF HUANGSHAN UNIVERSITY
年，卷(期)：2009，11(5)

参考文献(3条)

1. 刘冬喜.从约束型最小二乘估计到约束型岭估计的探索[J].湖南工业大学学报,2008,(3):40-42.

2. 赵天玺,李兆勤.回归系数的一种有偏估计[J].重庆工商大学学报(自然科学版),2007,(12):541-614.

3. 林路.回归系数的综合岭估计[J].数理统计与应用概率,1996,(3):179-184.

相似文献(10条)

提出递进自回归预测方法,其中包括递进自回归模型、递进自回归滑动平均模型、递进时变自回归模型、递进时变自回归滑动平均模型、递进回归-自回归模型.建立时间序列的递进预测公式,给出其最佳无偏预测,并推导出递进均方误差计算公式和高置信水平的递进预测区间估计.该方法是以逐步线性形式表示的一种非线性预测,既具有线性预测的简单性,又具有非线性预测精度高的特点.它不但可用于平稳时间序列预测,而且还可用于非平稳时间序列预测、确定性时间序列预测和小样本预测.此外,文中还给出时间序列线性组合及乘积的预测方法.并通过加权累加、倒数变换等方法,对观测值进行映射变换,使其呈现出更强的规律性,以进一步提高预测精度.