第 42 卷 第 7 期 西南师范大学学报(自然科学版) 2017 年 7 月
Vol. 42 No. 7 Journal of Southwest China Normal University (Natural Science Edition) Jul. 2017

DOI:10.13718/j.cnki.xsxb.2017.07.001

半参数可加模型参数的 Bayes 估计①

邬 吉 波

重庆市群与图的理论及应用重点实验室,重庆文理学院,重庆 永川 402160

摘要:研究了半参数可加模型参数的 Bayes 估计问题. 导出了半参数可加模型中参数的 Bayes 最小风险线性无偏估计,同时研究了其均方误差矩阵准则下优于 Profile 最小二乘估计的优良性.

关 键 词:半参数可加模型;Profile 最小二乘方法;均方误差矩阵准则

中图分类号:**O212.1** **文献标志码**:**A** **文章编号**:1000-5471(2017)07-0001-04

考虑如下的半参数可加模型:

$$Y_i = \boldsymbol{X}'_i\boldsymbol{\beta} + m_1(Z_{1i}) + \cdots + m_q(Z_{qi}) + \varepsilon_i,\ i = 1,2,\cdots,n \tag{1}$$

其中:Y_i 为因变量观察值;$\boldsymbol{X}_i$ 是 $p\times 1$ 维的解释变量;$\boldsymbol{\beta} = (\beta_1,\cdots,\beta_p)'$ 是未知的模型系数;$Z_{1i},\cdots,Z_{qi}$ 为自变量观察值;$m_1(\cdot),\cdots,m_q(\cdot)$ 为未知的光滑参数;ε_i 为独立的随机误差且 $E(\varepsilon_i) = 0$,$\mathrm{Cov}(\varepsilon_i) = \sigma^2$. 假设 $E(m_k(Z_k)) = 0$,$k = 1,\cdots,q$,这是为了保证模型的可识别性. 同时我们假定 Y_i 和 X_i 都是中心化的.

对于模型(1),很显然包括了一些常见的模型,当 $m_1(\cdot) = \cdots = m_q(\cdot) = 0$,模型(1) 就是经典的线性模型;当 $\boldsymbol{\beta} = \boldsymbol{0}$ 时,模型(1) 变成了可加模型;当 $q = 1$ 时,模型(1) 变成了半参数回归模型. 对于模型(1),我们比较关心参数部分的估计问题. 文献[1]提出了一种 Backfitting 算法,文献[2]提出了约束的 Profile 最小二乘估计,文献[3] 为了克服复共线性,提出了 Liu 估计.

对于模型(1) 中参数部分的 Bayes 估计目前还没有文献涉及,本文讨论参数部分的 Bayes 估计问题,导出了半参数可加模型中参数的 Bayes 最小风险线性无偏估计,同时讨论该估计的性质.

1 Profile 最小二乘估计

假设参数分量 $\boldsymbol{\beta}$ 已知,则可以把模型(1) 写成如下的可加模型

$$Y_i^* = Y_i - \boldsymbol{X}'_i\boldsymbol{\beta} = m_1(Z_{1i}) + \cdots + m_q(Z_{qi}) + \varepsilon_i,\ i = 1,2,\cdots,n \tag{2}$$

借鉴文献[4] 及文献[5] 中分析的 Backfitting 算法,定义 $\boldsymbol{M}$ 估计为

$$\hat{\boldsymbol{M}} = \boldsymbol{W_M}(\boldsymbol{Y} - \boldsymbol{X\beta}) \tag{3}$$

其中:$\boldsymbol{Y} = (Y_1,\cdots,Y_n)'$,$\boldsymbol{X} = (X'_1,\cdots,X'_n)'$,$M = \sum\limits_{k=1}^{q}\boldsymbol{m}_k$,$\boldsymbol{m}_k = [m_k(Z_{k1}),\cdots,m_k(Z_{kn})]'$,$\boldsymbol{W}_M$ 为模型(2) 对应的光滑矩阵,具体形式见文献[5].

如果我们将 $\hat{\boldsymbol{M}}$ 代替模型(1) 中的 $\boldsymbol{M}$,则可以得到如下的线性回归模型:

$$(\boldsymbol{I_n} - \boldsymbol{W_M})\boldsymbol{Y} = (\boldsymbol{I_n} - \boldsymbol{W_M})\boldsymbol{X\beta} + \varepsilon \tag{4}$$

记 $\bar{\boldsymbol{Y}} = (\boldsymbol{I_n} - \boldsymbol{W_M})\boldsymbol{Y}$,$\bar{\boldsymbol{X}} = (\boldsymbol{I_n} - \boldsymbol{W_M})\boldsymbol{X}$,则模型(4) 可以表示为:

① 收稿日期:2015-09-17

基金项目:国家自然科学基金青年项目(11501072,11426054);重庆市科学技术项目(cstc2015jcyjA00001);重庆市教委科学技术研究项目(KJ1501114).

作者简介:邬吉波(1985-),男,重庆垫江人,副教授,博士,主要从事线性模型、半参数模型的参数估计和变量选择的研究.

$$\bar{Y} = \bar{\boldsymbol{X}}\boldsymbol{\beta} + \varepsilon \tag{5}$$

对模型(5),采用最小二乘方法,得到参数的 Profile 最小二乘估计(PLS)

$$\hat{\boldsymbol{\beta}}_{PLS} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\bar{\boldsymbol{X}}'\bar{\boldsymbol{Y}} \tag{6}$$

2 参数分量的 Bayes 估计

本小节我们讨论参数分量的 Bayes 估计.

假设参数分量 $\boldsymbol{\beta}$ 的先验分布 $\pi(\boldsymbol{\beta})$ 满足条件

$$E(\boldsymbol{\beta}) = \mu,\ \mathrm{Cov}(\boldsymbol{\beta}) = \tau^2 I_p \tag{7}$$

这里 μ 和 $\tau^2 > 0$ 都是已知的常数.

对于模型(5),文献中求参数的 Bayes 估计主要有 2 种方法:一种是在正态分布的假设下假定 $\boldsymbol{\beta}$ 的先验分布为共轭先验,若为正态分布或者为无信息先验,则其在二次损失下 Bayes 估计由后验均值给出,详细内容参考文献[6];第二种方法是在 Gauss-Markov 模型下,假设 $\boldsymbol{\beta}$ 的先验分布的二阶矩存在,然后利用最优化的方法使线性估计的 Bayes 风险达到最小时确定 Bayes 估计,常常称为线性 Bayes 估计. 此方法是由文献[7] 提出的,文献[8-9] 讨论了此类方法. 由于第二种方法对样本分布和先验分布所加的条件较少,所以我们本文采用第二种方法.

假设 $\boldsymbol{\delta} = \boldsymbol{\delta}(\boldsymbol{x})$ 为参数向量 $\boldsymbol{\beta}$ 的一个估计. 本文定义损失函数为

$$L(\boldsymbol{\delta},\ \boldsymbol{\beta}) = (\boldsymbol{\delta} - \boldsymbol{\beta})'D(\boldsymbol{\delta} - \boldsymbol{\beta}) \tag{8}$$

这里 $\boldsymbol{D}$ 为已知正定矩阵,$R(\boldsymbol{\delta},\ \boldsymbol{\beta}) = E[L(\boldsymbol{\delta},\ \boldsymbol{\beta})]$ 为 $\boldsymbol{\delta}$ 的风险函数.

假定 $\boldsymbol{\beta}$ 的线性估计类为如下形式:

$$\tau = \{\hat{\boldsymbol{\beta}}^* = \boldsymbol{A}\bar{\boldsymbol{Y}} + \boldsymbol{b}\} \tag{9}$$

这里 $\boldsymbol{A}$ 和 $\boldsymbol{b}$ 分别是 $p \times n$ 和 $p \times 1$ 矩阵.

记 $\hat{\boldsymbol{\beta}}_{BE}$ 为参数分量 $\boldsymbol{\beta}$ 的 Bayesian 最小风险线性无偏估计,有定义 $\hat{\boldsymbol{\beta}}_{BE}$ 满足如下条件

$$R(\hat{\boldsymbol{\beta}}_{BE},\ \boldsymbol{\beta}) = \min_{A,b} R(\hat{\boldsymbol{\beta}}^*,\ \boldsymbol{\beta}) = \min_{A,b} E[(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})'D(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta})] \tag{10}$$

同时满足如下约束条件 $E(\hat{\boldsymbol{\beta}}^* - \boldsymbol{\beta}) = \boldsymbol{0}$. 由此约束条件可以求得:

$$b = (\boldsymbol{I}_n - \boldsymbol{A}\bar{\boldsymbol{X}})\boldsymbol{\mu} \tag{11}$$

为了求得 $\boldsymbol{A}$ 的值,我们首先计算 Bayes 风险

$$\begin{aligned} R(\hat{\boldsymbol{\beta}}^*,\ \boldsymbol{\beta}) = & E\{[\boldsymbol{A}(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}\boldsymbol{\mu}) - (\boldsymbol{\beta} - \boldsymbol{\mu})]'\boldsymbol{D}[\boldsymbol{A}(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}\boldsymbol{\mu}) - (\boldsymbol{\beta} - \boldsymbol{\mu})]\} = \\ & E\{tr\{\boldsymbol{D}[\boldsymbol{A}(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}\boldsymbol{\mu}) - (\boldsymbol{\beta} - \boldsymbol{\mu})][\boldsymbol{A}(\bar{\boldsymbol{Y}} - \bar{\boldsymbol{X}}\boldsymbol{\mu}) - (\boldsymbol{\beta} - \boldsymbol{\mu})]'\}\} = \\ & \sigma^2 tr[\boldsymbol{D}\boldsymbol{A}(\boldsymbol{I}_n + \rho^{-1}\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}')\boldsymbol{A}' + \rho^{-1}\boldsymbol{D} - \rho^{-1}\boldsymbol{D}\boldsymbol{A}\bar{\boldsymbol{X}} - \rho^{-1}\boldsymbol{D}\boldsymbol{A}\bar{\boldsymbol{X}}'\boldsymbol{A}'] \end{aligned} \tag{12}$$

这里 $\rho = \frac{\sigma^2}{\tau^2}$. 求 $R(\hat{\boldsymbol{\beta}}^*,\ \boldsymbol{\beta})$ 得最小值,可以得到 $\boldsymbol{A}$ 的值. 由矩阵微商准则以及令 $\frac{\partial R(\hat{\boldsymbol{\beta}}^*,\ \boldsymbol{\beta})}{\partial A} = 0$,得到

$$\boldsymbol{D}\boldsymbol{A}(\boldsymbol{I}_n + \rho^{-1}\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}') - \rho^{-1}\boldsymbol{D}\bar{\boldsymbol{X}}' = \boldsymbol{0} \tag{13}$$

解得

$$\boldsymbol{A} = \rho^{-1}\bar{\boldsymbol{X}}'(\boldsymbol{I}_n + \rho^{-1}\bar{\boldsymbol{X}}\bar{\boldsymbol{X}}')^{-1} \tag{14}$$

利用文献[10],可将 $\boldsymbol{A}$ 写成

$$\boldsymbol{A} = (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} + \boldsymbol{I}_p)^{-1}\bar{\boldsymbol{X}}' \tag{15}$$

所以参数向量 $\boldsymbol{\beta}$ 的 Bayes 估计为

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{BE} = & (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} + \boldsymbol{I}_p)^{-1}(\bar{\boldsymbol{X}}'\bar{\boldsymbol{Y}} + \rho\boldsymbol{\mu}) = \\ & (\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}} + \boldsymbol{I}_p)^{-1}((\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}\hat{\boldsymbol{\beta}}_{PLS} + \rho\boldsymbol{\mu}) = \end{aligned}$$

$$\hat{\boldsymbol{\beta}}_{PLS}-(\rho\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}+\boldsymbol{I}_p)^{-1}(\hat{\boldsymbol{\beta}}_{PLS}-\boldsymbol{\mu}) \tag{16}$$

即式(16)为参数向量$\boldsymbol{\beta}$的Bayes估计．在下节我们在均方误差矩阵准则下讨论Bayes估计$\hat{\boldsymbol{\beta}}_{BE}$相对于Profile最小二乘估计的优良性.

3　均方误差矩阵准则下 Bayes 估计的优良性

为了得到$\hat{\boldsymbol{\beta}}_{BE}$的优良性，首先我们给出如下定义.

定义 1　假设$\hat{\boldsymbol{\theta}}$是参数向量$\boldsymbol{\theta}$的一个估计，则$\hat{\boldsymbol{\theta}}$的均方误差(MSE)定义为

$$\mathrm{MSE}(\hat{\boldsymbol{\theta}})=E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})'(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})]$$

$\hat{\boldsymbol{\theta}}$的均方误差矩阵(MSEM)定义为

$$\mathrm{MSEM}(\hat{\boldsymbol{\theta}})=E[(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})']$$

假设$\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2$为参数$\boldsymbol{\theta}$的2个不同估计，如果$\mathrm{MSEM}(\hat{\boldsymbol{\theta}}_2)-\mathrm{MSEM}(\hat{\boldsymbol{\theta}}_1)\geqslant 0$(或$\mathrm{MSE}(\hat{\boldsymbol{\theta}}_2)-\mathrm{MSE}(\hat{\boldsymbol{\theta}}_1)\geqslant 0$)，则称在MSEM(或MSE)准则下，$\hat{\boldsymbol{\theta}}_1$优于$\hat{\boldsymbol{\theta}}_2$.

下面我们给出Bayes估计$\hat{\boldsymbol{\beta}}_{BE}$和Profile最小二乘估计的比较结果.

定理 1　假设模型(1)中参数向量$\boldsymbol{\beta}$的Bayes估计和Profile最小二乘估计分别由式(16)和式(6)给出，则

1) $\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{PLS})-\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{BE})>0$

2) $\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{PLS})-\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{BE})>0$

证　先证明1)，由式(6)以及定义(1)可得：

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{PLS})=E[(\hat{\boldsymbol{\beta}}_{PLS}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{PLS}-\boldsymbol{\beta})']=\sigma^2(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1} \tag{17}$$

同理由式(16)和定义(1)可得

$$\begin{aligned}\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{BE})=&E[(\hat{\boldsymbol{\beta}}_{BE}-\boldsymbol{\beta})(\hat{\boldsymbol{\beta}}_{BE}-\boldsymbol{\beta})']=\\&\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{BE}-\boldsymbol{\beta})=\\&E[\mathrm{Cov}[(\hat{\boldsymbol{\beta}}_{BE}-\boldsymbol{\beta})\mid\boldsymbol{\beta}]]+\mathrm{Cov}[E[(\hat{\boldsymbol{\beta}}_{BE}-\boldsymbol{\beta})\mid\boldsymbol{\beta}]]=\\&\sigma^2(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}+\rho\boldsymbol{I}_p)^{-1}\end{aligned} \tag{18}$$

所以我们有：

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{PLS})-\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{BE})=\sigma^2[(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}})^{-1}-(\bar{\boldsymbol{X}}'\bar{\boldsymbol{X}}+\rho\boldsymbol{I}_p)^{-1}] \tag{19}$$

由文献[10]的结论：如果矩阵$\boldsymbol{A}>\boldsymbol{B}>\boldsymbol{0}$，则$\boldsymbol{B}^{-1}-\boldsymbol{A}^{-1}>\boldsymbol{0}$. 所以

$$\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{PLS})-\mathrm{MSEM}(\hat{\boldsymbol{\beta}}_{BE})>0$$

由于$\mathrm{MSE}(\hat{\boldsymbol{\theta}})=tr[\mathrm{MSEM}(\hat{\boldsymbol{\theta}})]$，易得$\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{PLS})-\mathrm{MSE}(\hat{\boldsymbol{\beta}}_{BE})>0$.

参考文献：

[1]　OPSOMER J D，RUPPERT D. A Root-n Consistent Backfitting Estimator for Semiparametric Additive Modeling [J]. Journal of Computational and Graphical Statistics，1999，8(4)：715－732.

[2]　WEI C H，LIU C. Statistical Inference on Semi-Parametric Partial Linear Additive Models [J]. Journal of Nonparametric Statistics，2012，24(4)：809－823.

[3]　王肖南，魏传华. 参数可加模型的 Liu 估计[J]，统计与决策，2015(8)：28－30.

[4]　BIJA A，HASTIET，TIBSHIRANI R. Linear smoothers and Additive Models(with Discussion) [J]. Annals of Statistics，1989，17：453－555.

[5] OPSOMER J D. Asymptotic Properties of Backfitting Estimators [J]. Journal of Multivariate Analysis. 2000, 73(2): 5-7.
[6] WANG S G, CHOW S C. Advanced Linear Models Theory and Applications [M]. New York: Marcel Dekker, 1994.
[7] RAO C R. Linear Statistical Inference and It Applications [M]. 2nd ed. New York: Wiley, 1973.
[8] WEI L S, ZHANG W P. The Superiorities of Bayes Linear Minimum Risk Estimation in Linear Model [J]. Communications in Statistics: Theory and Methods, 2007, 36: 917-926.
[9] 李琪琪,韦来生. 半参数回归模型中参数的 Bayes 估计 [J]. 中国科学技术大学学报, 2010, 40(9): 881-886.
[10] RAO C R, TOUTENBUR, H, SHALABH, et al. Linear Models and Generalizations, Least Squares and Alternatives [M]. Berlin: Springer, 2008.

The Bayes Estimation of Parameters in Semiparametric Partially Linear Additive Model

WU Ji-bo

Key Laboratory of Group & Graph Theories and Applications, Chongqing University of Arts and Sciences, Yongchuan Chongqing 402160, China

Abstract: In this paper we study Bayes estimator of parameters in semiparametric partially linear additive model. The Bayes minimum risk linear unbiased estimator of parameters was derived in the semiparametric partially linear additive model. The superiority of the new estimator over Profile least squares estimator was discussed in terms of mean square error matrix criterion.

Key words: semiparametric partially additive model; profile least squares method; mean square error matrix criterion

责任编辑 张 栒