应用概率统计 第二十七卷
第一期 2011年2月

Chinese Journal of Applied Probability
and Statistics Vol.27 No.1 Feb. 2011

含高维相依自变量的中心k阶条件矩子空间的估计 *

徐 群 芳

(浙江农林大学统计系, 临安, 311300)

摘 要

在回归分析中往往对条件均值, 条件方差及高阶条件矩特别感兴趣. 本文我们将关注中心k阶条件矩子空间在高维相依自变量情形的估计问题. 为此, 我们首先引入中心k阶条件矩子空间的概念, 并研究该子空间的基本性质. 针对高维相依自变量的复杂数据, 为了避免预测变量协方差阵的逆矩阵的计算, 本文提出用偏最小二乘方法来估计中心k阶条件矩子空间. 最后得到了估计的强相合性等渐近性质.

关键词: 充分降维子空间, 中心k阶条件矩子空间, 高维相依, 最小二乘估计, 偏最小二乘.

学科分类号: O212.1.

§1. 引 言

现代科学技术和社会经济的许多领域都会遇到高维相依自变量复杂数据的统计分析问题. 它是目前统计学应用和理论中备受关注的突出问题. 一方面由于高维数据分析十分困难, 另一方面高维数据中的信息往往主要包含在一个或几个低维结构中, 因此降维是分析高维相依数据的一个重要手段.

一般地, 对于一维响应变量Y和p维预测变量$X=(X_1,\cdots,X_p)^T$之间的回归问题, 主要目标是要推断Y在给定X下的条件分布$F_{Y|X}$. 即研究对于X的不同取值, $Y|X$如何变化? 当解释变量X的维数p较大时, 充分降维(sufficient dimension reduction, 简称SDR)理论提供了回归分析的一个有效出发点. 充分降维的基本目标就是要寻找估计p维解释变量X的最小线性组合$\eta_1^TX,\cdots,\eta_q^TX$, $q\leq p$, 令$p\times q$的矩阵$\eta=(\eta_1,\cdots,\eta_q)$, 使得用更低维的$\eta^TX$来代替$X$, 而不损失从$X$中可获得的条件分布$F_{Y|X}$的回归信息, 即包含所有从样本中可获得的回归信息, 而且不需要指定参数模型. 充分降维是要寻找一个中心降维子空间, 也称中心子空间(central subspace, 简称CS), 记为$\mathcal{S}_{Y|X}$. 它是所有满足给定η^TX时Y与X条件独立(即对所有可能的X值, 条件分布$Y|X$与$Y|\eta^TX$服从相同的分布)的子空间的交集. 若存在, 就是最小子空间. 在很弱的条件下CS总是存在的, 矩阵η的列向量构成中心子空间的一组基, η^TX包含了关于条件分布$F_{Y|X}$的所有回归信息. 因此，若中心子空间$\mathcal{S}_{Y|X}$已知，则可用Y关于η^TX的最小充分回归概要图来进行系列回归分析. 若$\widehat{\eta}$是中心子空间$\mathcal{S}_{Y|X}$的基的估计, 则可用Y关于$\widehat{\eta}^TX$的最小充分回归概要图来近似(Cook (1998), Cook和Weisberg (1999)).

*浙江省教育厅项目(20070939)资助.

本文2009年10月12日收到, 2010年4月18日收到修改稿.

中心子空间CS是包含条件分布$F_{Y|X}$的所有回归信息的最小子空间, 刻画了Y与X的完全相依关系. 但当回归分析特别关注推断条件均值$\mathsf{E}(Y|X)$, 而把条件分布$F_{Y|X}$的其它信息作为讨厌参数时, Cook和Li (2002)引出了一个新的降维子空间, 中心均值子空间(central mean subspace, 简称CMS), 并记为$\mathcal{S}_{\mathsf{E}(Y|X)}$. 中心均值子空间(CMS)的定义与中心子空间(CS)类似, 它是所有满足给定A^TX时Y与$\mathsf{E}(Y|X)$条件独立的子空间的交集. 在很弱的条件下CMS也总是存在的, 矩阵A的列向量构成中心均值子空间$\mathcal{S}_{\mathsf{E}(Y|X)}$的一组基, A^TX包含了关于条件均值$\mathsf{E}(Y|X)$的所有回归信息. 易见$\mathcal{S}_{\mathsf{E}(Y|X)}\subseteq\mathcal{S}_{Y|X}$. Yin和Cook (2002)进一步推广了中心均值子空间的概念, 提出了中心k阶矩降维子空间.

近年来有不少文献研究估计中心子空间CS和中心均值子空间CMS的方法, 如最小二乘(ordinary least squares, 简称OLS), 切片逆回归(sliced inversion regression, 简称SIR), 部分切片逆回归(partial SIR), 切片均方差估计(sliced average variance estimation, 简称SAVE), 主黑塞方向(principal Hessian direction, 简称PHD)以及方向回归(directional regression, 简称DR)等. 但是这些方法都没考虑预测变量的维数依赖于样本量的情形, 而Zhu和Zhu (2009)针对该情形, 提出了用分布加权的偏最小二乘方法估计中心子空间, 并证明了估计具有相合性和渐近正态性等大样本性质.

本文我们将进一步推广中心均值子空间, 特别关注k阶条件矩$\mathsf{E}(Y^k|X)$的回归推断问题. 接下来, 我们在论文第二部分将引入中心k阶条件矩子空间的概念, 并研究该子空间的性质; 在第三部分给出高维相依自变量的中心k阶条件矩子空间的基于k阶矩加权偏最小二乘估计方法以及相合性等结论; 模拟和讨论分别在第四第五部分给出; 为了论文主线的简明, 把较复杂的定理证明放在论文最后.

§2. 基本假设及相关引理

全文假定: 一维响应变量Y和p维预测变量$X=(X_1,\cdots,X_p)^T$具有联合分布且具有有限矩, $F_{Y|X}$为Y在给定X下的条件分布. 样本$\{(x_i,y_i),i=1,\cdots,n\}$是来自总体$(X,Y)$的独立同分布样本. 为讨论方便, 约定预测变量是中心化的, 满足$\mathsf{E}(X)=0$和$\mathsf{Cov}\,(X)=\Sigma$, 除非特别说明. 引入论文中一些记号: $U\perp\!\!\!\perp V|Z$是指在给定随机向量Z的任何一个取值时, 随机向量U和V条件独立. 子空间记为$\mathcal{S}$, $\mathcal{S}_B=\mathrm{span}\{B\}$表示由矩阵$B$的列向量所张成的子空间. P_B表示在标准内积下到子空间$\mathcal{S}_B$的投影算子, 而$P_B(\Sigma)$表示在Σ内积下到子空间$\mathcal{S}_B$的投影算子. 另外, 现有子空间的一些估计方法需要以下两个假设条件(Cook和Li (1999)):

条件 1 (线性条件) 条件均值$\mathsf{E}(X|B^TX)$关于X是线性的, 即

$$\mathsf{E}(X|B^TX)=\Sigma B(B^T\Sigma B)^{-1}B^TX,\qquad \Sigma=\mathsf{Cov}\,(X). \tag{2.1}$$

条件 2 (常数方差条件) 条件方差$\mathsf{Var}\,(X|B^TX)$与Y不相关.

中心子空间$\mathcal{S}_{Y|X}$是包含条件分布$F_{Y|X}$的所有回归信息的最小子空间, 刻画了Y与X的完全相依关系. 在Cook和Li (2002)引出的中心均值子空间$\mathcal{S}_{\mathsf{E}(Y|X)}$包含了关于条件均值$\mathsf{E}(Y|X)$的所有回归信息, 而把条件分布$F_{Y|X}$的其它信息作为讨厌参数. 下面我们将特

别关注k阶条件矩$\mathsf{E}(Y^k|X)$, 而把条件分布$F_{Y|X}$的其它信息作为讨厌参数的回归推断问题. Yin和Cook (2002)进一步推广了中心均值子空间的概念, 提出了中心k阶矩降维子空间$\mathcal{S}_{Y|X}^{(k)}$. 它包含了关于前k阶条件中心矩$\mathsf{E}(Y^j|X)$, $j=1,\cdots,k$的所有回归信息. 该文献中提出的协方差估计方法要用到Σ^{-1}的估计, 沿着类似的思路, 我们先考虑其特殊情形, 引入中心k阶条件矩子空间的概念, 并研究该子空间的性质.

定义 2.1 令B为满足以下条件的$p\times q\,(q\leq p)$矩阵,

$$Y \perp\!\!\!\perp \mathsf{E}(Y^k|X)|B^TX, \tag{2.2}$$

则$\mathcal{S}_B$是一个Y关于X的k阶条件矩DRS. 令$\mathcal{S}_{\mathsf{E}(Y^k|X)}$是所有$k$阶条件矩DRSs的交集, 若$\mathcal{S}_{\mathsf{E}(Y^k|X)}$也是$k$阶条件矩DRS, 则称其为中心$k$阶条件矩子空间(central kth-conditional moment subspace, 简称CKCMS), 记为$\mathcal{S}_{\mathsf{E}(Y^k|X)}$.

中心k阶条件矩子空间CKCMS不一定存在, 但在微弱的条件小就能保证其存在. 例如, 若X的支撑是开的凸集, 则CS, CMS和CKCMS都存在. 由于存在性不是很关键的问题, 故在论文接下来部分, 我们总假定子空间是存在的. 中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$是最小的$k$阶条件矩降维子空间, 矩阵$B$的列向量构成$\mathcal{S}_{\mathsf{E}(Y^k|X)}$的一组基, B^TX包含了关于k阶条件均值$\mathsf{E}(Y^k|X)$的所有回归信息. 显见, $\mathcal{S}_{\mathsf{E}(Y|X)}\subseteq\mathcal{S}_{\mathsf{E}(Y^k|X),k=1,\cdots,j}\subseteq\mathcal{S}_{Y|X}$.

下面的引理给出了定义2.1中(2.2)条件独立的几个等价条件.

引理 2.1 下面的命题是等价的.

(1) $Y \perp\!\!\!\perp \mathsf{E}(Y^k|X)|B^TX$, $k=1,\cdots,j$.

(2) $\mathsf{Cov}\,\{Y^k,\mathsf{E}(Y^k|X)|B^TX\}=0$, $k=1,\cdots,j$.

(3) $\mathsf{E}(Y^k|X)$是B^TX的函数, $k=1,\cdots,j$.

(4) 对任意函数$f(\cdot)$, $\mathsf{Cov}\,\{Y^k,f(X)|B^TX\}=0$, $k=1,\cdots,j$.

该引理是Yin和Cook (2002)命题1的特殊情形.

§3. 子空间CKCMS的估计方法及结论

在这一节我们给出寻找高维相依预测变量的中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$方向的估计方法, 基于$k$阶矩加权偏最小二乘以及相合性等结论.

3.1 基于k阶矩加权最小二乘估计

在回归分析中, 最小二乘法(简称OLS)是简单易实现的估计方法. Cook和Li (2002)指出: 在线性条件(2.1)下, 利用OLS可以寻找估计中心均值子空间CMS的方向, $\beta_{\rm OLS}=\Sigma^{-1}V$, 其中$\Sigma=\mathsf{Cov}\,(X)$, $V=\mathsf{Cov}\,(X,Y)$, 而且有${\rm span}\{\beta_{\rm OLS}\}\subseteq\mathcal{S}_{\mathsf{E}(Y|X)}$.

类似的, 对于中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$方向的估计, 也可以用OLS想法. 令

$$\beta_{k0}=\mathsf{Cov}\,(X,Y^k)=\mathsf{E}(XY^k),\qquad k=1,\cdots,j. \tag{3.1}$$

上述式子中最后一个等式是由于假定$\mathsf{E}(X)=0$. 类似于Yin和Cook (2002)命题2, 我们可以证明在很弱的条件下$\Sigma^{-1}\beta_{k0}$在中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$上, 得到的结论如下:

引理 3.1 对任意固定的k, 假设矩阵B的列向量构成中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$的一组基, 即$\mathcal{S}_B=\mathrm{span}\{B\}=\mathcal{S}_{\mathsf{E}(Y^k|X)}$. 则在线性条件(2.1)下, 有

$$\beta_{k0}=\mathsf{E}(XY^k)\in\Sigma\mathcal{S}_{\mathsf{E}(Y^k|X)}. \tag{3.2}$$

显见引理3.1的结论等价于$\mathrm{span}\{\beta_{k0}\}\subseteq\Sigma\mathcal{S}_{\mathsf{E}(Y^k|X)}\subseteq\Sigma\mathcal{S}_{Y|X}$. 也即

$$\Sigma^{-1}\beta_{k0}=\Sigma^{-1}\mathsf{E}(XY^k)\in\mathcal{S}_{\mathsf{E}(Y^k|X)}\subseteq\mathcal{S}_{Y|X}.$$

不难看出, 与估计CMS空间方向的OLS估计$\beta_{\mathrm{OLS}}=\Sigma^{-1}\mathsf{Cov}\,(X,Y)=\Sigma^{-1}\mathsf{E}(XY)$相比较, 我们提出的CKCMS的方向的估计$\Sigma^{-1}\beta_{k0}=\Sigma^{-1}\mathsf{E}(XY^k)$是用$\mathsf{E}(XY^k)$代替$\mathsf{E}(XY)$. 为此, 我们把估计$\beta_{\mathrm{KMWLSE}}\doteq\Sigma^{-1}\beta_{k0}=\Sigma^{-1}\mathsf{E}(XY^k)$称为基于$k$阶矩加权最小二乘估计.

3.2 基于k阶矩加权偏最小二乘

由引理3.1的结论可知, 可用$\Sigma^{-1}\beta_{k0}=\Sigma^{-1}\mathsf{E}(XY^k)$来寻找估计中心$k$阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$的方向, 从而也有助于推断中心子空间CS. 但是从中可见, 估计时涉及预测变量协方差阵$\Sigma=\mathsf{Cov}\,(X)$的逆矩阵的计算和估计. 而当预测变量之间存在高维严重多重相关性的复杂数据或解释变量多而样本量相对较少的条件下, Σ往往是病态的或不可逆的, 普通最小二乘法往往失效, 不仅会增大模型误差, 而且使模型丧失稳健性, 而偏最小二乘方法(partial least squares, 简称PLS)的优越性便能充分显示出来. 偏最小二乘法是一种新型的处理高维共线性预测变量的多元统计数据分析方法, 它于1983年由伍德(S. Wold)和阿巴诺(C. Albano)等人首次提出. PLS的基本思想可详细参见文献Wold (1966, 1975), Wold, Martens和Wold (1983)等. 近几十年来, PLS在理论、方法和应用方面都得到了迅速的发展, 其中Naik和Tsai (2000)证明了在很弱的条件下, 用$V=\mathsf{Cov}\,(X,Y)$作为种子向量的PLS估计存在于中心均值子空间$\mathcal{S}_{\mathsf{E}(Y|X)}$中. 我们感兴趣的是如何把该方法进行适当的改进推广, 以寻求用PLS方法来估计中心k阶条件矩子空间CKCMS. 下面我们给出用基于k阶矩加权偏最小二乘方法来估计中心k阶条件矩子空间CKCMS的方向. 为了避免使用Σ^{-1}, 我们首先假定存在子空间$\mathrm{span}\{R\}$包含$\Sigma^{-1}\beta_{k0}$. 然后我们把$\Sigma^{-1}\beta_{k0}$在Σ内积下投影到子空间$\mathrm{span}\{R\}$上, 即有$\Sigma^{-1}\beta_{k0}=R(R^T\Sigma R)^{-1}R^T\beta_{k0}$. 这样, 就不需要计算和估计$\Sigma^{-1}$了. 有关基矩阵$\{R\}$的选择的基本原则是在保证不包含$\Sigma^{-1}$的前提下不能太大, 也不能过小. 为此, 我们根据PLS方法的思想, 建议用$\beta_{k0}=\mathsf{E}(XY^k)$作为种子向量, 把$\Sigma^{-1}\beta_{k0}$在$\Sigma$内积下投影到如下Krylov子空间

$$R_u=(\beta_{k0},\Sigma\beta_{k0},\cdots,\Sigma^{u-1}\beta_{k0}),\qquad u=1,2,\cdots, \tag{3.3}$$

于是, 我们可以得到该投影为

$$R_u(R_u^T\Sigma R_u)^{-1}R_u^T\beta_{k0}=:\beta_{ku},\qquad u=1,2,\cdots. \tag{3.4}$$

由于当u增加时, 在式(3.3)中的$p \times u$矩阵R_u张成非减嵌套子空间序列. 故当$\Sigma^{-1}\beta_{k0} \in \mathrm{span}\{R_u\}$时, 我们可以得到

$$\Sigma^{-1}\beta_{k0} = \beta_{ku}. \tag{3.5}$$

在此论文中, 我们把由式(3.4)定义的并满足式(3.5)的β_{ku}称为基于k阶矩加权偏最小二乘(the kth-moment weight partial least squares, 简称KMWPLS).

显然, 计算(3.4)给出的β_{ku}不包含Σ的逆矩阵. 为了研究β_{ku}的性质和u的合理选择, 我们给出如下条件, 类似于Naik和Tsai (2000)中所用的条件.

条件 3　$\beta_{k0} = \sum\limits_{i=1}^{m} \theta_i\gamma_i$, 其中$\theta_1, \cdots, \theta_m$是非零实数, $\gamma_1, \cdots, \gamma_m$是$\Sigma$的$m$个不同的正的特征值对应的特征向量.

注意到假如Σ的特征值是多重的, 则我们可以再采取Helland's (1990)所提出的方法. 特别, 若Σ是与$p \times p$单位矩阵成比例的矩阵, 则可取$m = 1$以及$\gamma_1 = \beta_{ku}$. 下面我们给出β_{ku}的一个性质.

定理 3.1　假定模型满足条件1 (线性条件)和条件3以及$u = m$. 则

$$\beta_{ku} = \beta_{km} \subseteq \mathcal{S}_{\mathsf{E}(Y^k|X)}.$$

由于$\beta_{km} = R_m(R_m^T\Sigma R_m)^{-1}R_m^T\beta_{k0}$不涉及$\Sigma$的逆矩阵, 故可用它来估计高维严重相依预测变量的中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$的方向.

注记 1　关于u的合理选择, Zhu和Zhu (2009)指出, 当条件1满足时, u的最优选择为Σ的不同的正的特征值个数m, 而且$u = m = \dim(R_p) = \dim(R_pR_p^T)$. 又由于$m$相对于维数$p$来说很小, 为了进一步提高具体实现时估计的效率, 我们可以通过选择满足$p \gg K_n \geq m$的合适的K_n. 显然, 矩阵$R_pR_p^T$和$R_{K_n}R_{K_n}^T$的维数相等都是m. 因此我们只需要计算更低维的矩阵R_{K_n}的维数来估计m. 矩阵$R_{K_n}R_{K_n}^T$的非零特征值个数可以作为$u = m$的合理选择.

注记 2　由方程(3.3)、(3.4)和(3.5), 我们可得到优化初始向量β_{k0}的迭代算法以提高具体实现时估计的准确性. 令$\beta_{k0}^{(0)} = \mathsf{E}(XY^k)$是初始值. $\beta_{k0}^{(i)}$的第i步迭代公式为

$$\beta_{k0}^{(i)} = \Sigma R_u^{(i-1)}(R_u^{(i-1),T}\Sigma R_u^{(i-1)})^{-1}R_u^{(i-1),T}\beta_{k0}^{(i-1)},$$

其中$R_u^{(i-1)} = (\beta_{k0}^{(i-1)}, \Sigma\beta_{k0}^{(i-1)}, \cdots, \Sigma^{u-1}\beta_{k0}^{(i-1)})$. 反复使用如上迭代公式, 直至$\beta_{k0}^{(i)}$不再变化, 满足所需要的精度要求. 在记号不至于混淆的情况下, 我们记β_{k0}的最终估计为

$$\beta_{k0} = \Sigma R_u^{(i)}(R_u^{(i),T}\Sigma R_u^{(i)})^{-1}R_u^{(i),T}\beta_{k0}^{(i)}.$$

3.3　基于k阶矩加权偏最小二乘估计的渐近性质

下面我们研究基于k阶矩加权偏最小二乘(KMWPLS)估计的渐进性质. 我们用Zhu和Zhu (2009)所提出的改进的贝叶斯信息准则(Bayesian information criterion, 简称BIC-型准

则)来估计矩阵$R_{K_n}R_{K_n}^T$的维数m, 并讨论了KMWPLS估计的相合性. 至于KMWPLS估计的渐近正态性及证明还有待以后进一步讨论.

由样本点数据$\{(x_i,y_i),i=1,\cdots,n\}$, 我们先用以下的矩估计来估计$\beta_{k0}$和$\Sigma$:

$$\widehat{\beta}_{k0}=\frac{1}{n}\sum_{i=1}^n x_iy_i^k,\qquad \widehat{\Sigma}=\frac{1}{n}\sum_{i=1}^n x_ix_i^T-\frac{1}{n^2}\sum_{i=1}^n x_i\sum_{i=1}^n x_i^T. \tag{3.6}$$

于是我们就可以得到对应估计的嵌套子空间序列$\{\widehat{R}_u,u=1,\cdots,K_n\}$, 其中当$n\to\infty$时$K_n\to\infty$. 例如我们可取$K_n=O((\log p)^{3/4})$, 当$p$大时, 它大于$m$. 由前面注记1的讨论, 矩阵$R_{K_n}R_{K_n}^T$的非零特征值个数可以作为$u=m$的合理选择. 因此, 我们可用Zhu和Zhu (2009)所提出的如下改进的BIC-型准则来估计矩阵$R_{K_n}R_{K_n}^T$的维数m:

令$\widehat{\lambda}_i$是矩阵$\widehat{R}_{K_n}\widehat{R}_{K_n}^T$的第$i$个特征值.

$$G(j)=\frac{\sum_{i=1}^{j}[\log(\widehat{\lambda}_i+1)-\widehat{\lambda}_i]}{\sum_{i=1}^{K_n}[\log(\widehat{\lambda}_i+1)-\widehat{\lambda}_i]}-C_n\times\frac{j(j+1)}{2K_n^2},\qquad j=1,\cdots,K_n.$$

从而m的估计定义为$G(\widehat{m}_k)=\max\limits_{1\le j\le K_n}G(j)$.

Zhu和Zhu (2009)定理1证明了对于$p=o(n^{1/2})$情形, 若令

$$C_n=O(\log n),\qquad K_n=O((\log p)^{3/4}),$$

则所得到的估计$\widehat{m}_k$是强相合的.

于是, 基于如上m的相合估计, 我们可以通过$\widehat{R}_{\widehat{m}_k}=(\widehat{\beta}_{k0},\widehat{\Sigma}\widehat{\beta}_{k0},\cdots,\widehat{\Sigma}^{\widehat{m}_k-1}\widehat{\beta}_{k0})$来估计Krylov子空间. 因此, 最终的基于$k$阶矩加权偏最小二乘(KMWPLS)的估计为

$$\widehat{\beta}_{k\widehat{m}_k}=\widehat{R}_{\widehat{m}_k}(\widehat{R}_{\widehat{m}_k}^T\widehat{\Sigma}\widehat{R}_{\widehat{m}_k})^{-1}\widehat{R}_{\widehat{m}_k}^T\widehat{\beta}_{k0}. \tag{3.7}$$

记号$\|\cdot\|$表示矩阵$\|A\|$的欧氏范数, 也称Frobenius范数, 即$\|A\|$是等于矩阵A的所有元素的平方之和的平方根. 关于KMWPLS的估计$\widehat{\beta}_{k\widehat{m}_k}$的渐近性质, 我们有如下相合性结果.

定理 3.2 假定$\widehat{m}_k\to m$几乎处处满足, 而且$\max\limits_{1\le i\le p}\mathsf{E}(X_i^4)<\infty$对$p$一致成立, 则

$$\|\widehat{\beta}_{k\widehat{m}_k}-\beta_{km}\|=o(pn^{-1/2}\log n),\qquad \text{几乎处处成立.}$$

§4. 模　拟

为了考察本文所提出的高维相依中心k阶条件矩子空间的方向的KMWPLS估计的效果, 我们用KMWPLS估计得到的估计空间span$\{\widehat{\beta}\}$和真实空间span$\{\beta\}$之间的绝对相关系数$|\mathrm{Corr}(\widehat{\beta}^TX,\beta^TX)|$的大小来衡量估计效果. 显然, 绝对相关系数值越大, 估计效果越好. 考虑以下模型:

$$Y=(x^T\beta-2)^2+\varepsilon,\qquad X\sim 0.5\times N_p(0,\Sigma),\qquad \varepsilon\sim 0.5\times N(0,1).$$

假设ε和X相互独立. 模型的参数选择如下: 解释变量X的协方差阵Σ是主对角线元素都是1其他元素都是0.8的p阶矩阵, 即解释变量之间是高度相关的. 令p维方向向量$\beta=(1,\cdots,1,0,\cdots,0)^T/\sqrt{5}$, 即前5个分量都是$1/\sqrt{5}$其他分量都是0. 我们从该模型中每次产生$n$个数据, 对不同的$n$和$p$分别都进行了100次重复模拟试验. 常见情形$k=1$的模拟结果见表1.

表1 绝对相关系数的均值和标准差(重复100次)

Normal predictors, and all are equally correlated.						
n	$p=10$	$p=20$	$p=30$	$p=40$	$p=50$	$p=100$
50	0.9554	0.9333	0.9102	0.8888	0.8836	0.8536
	0.0232	0.0257	0.0292	0.0312	0.0314	0.0372
100	0.9735	0.9510	0.9374	0.9205	0.9064	0.8669
	0.0122	0.0181	0.0228	0.0261	0.0319	0.0338
200	0.9849	0.9692	0.9621	0.9508	0.9392	0.9048
	0.0079	0.0117	0.0123	0.0130	0.0174	0.0205
400	0.9921	0.9841	0.9759	0.9700	0.9621	0.9397
	0.0044	0.0050	0.0077	0.0077	0.0100	0.0136

(每个格子中第一个是均值, 第二个是标准差.)

在表1中, 我们列出了对不同的n和p分别都进行了100次重复模拟得到的绝对相关系数的均值和标准差. 从表1的模拟结果, 我们可以看到对于不同的维数p, 即使较大的维数p的高维相依情形, 绝对相关系数的均值都较大而标准差都较小, 而且随着样本量n的增加, 绝对相关系数的均值越大而标准差越小. 可见该估计有很好的相合性的有限样本的具体实现. 而且对较小的样本量n和相对较大的维数p, 该估计仍有较好的估计效果.

§5. 讨 论

在本论文中, 我们把最小二乘估计和偏最小二乘方法进行改进, 推广到估计高维相依中心k阶条件矩子空间$\mathcal{S}_{\mathsf{E}(Y^k|X)}$的方向, 提出了基于$k$阶矩加权偏最小二乘估计. 该方法有以下主要优点:

(1) 在统计分析中, 往往对比条件分布简单的条件均值, 条件方差及高阶条件矩特别感兴趣.

(2) 已有的SIR, PHD和SAVE等方法虽都是较有效地降维方法, 但这些方法都基于下面的两个假设: 线性条件和常数方差条件. 而且这些方法都要用到预测变量的协方差阵Σ的逆矩阵的计算和估计. 而本文给出的基于k阶矩加权偏最小二乘估计方法只需要其中的线性条件, 而且避免了协方差阵Σ的逆矩阵的计算和估计. 适用于高维相依预测向量情形.

另外, 在高维相依预测变量情形, 利用KMWPLS方法得到各中心$k\,(k=1,2,\cdots,j)$阶条件矩子空间方向的估计后, 易求得中心前j阶条件矩子空间的估计. 中心前j阶条件

矩子空间类似于中心均值子空间的定义. 令B为满足以下条件的$p\times q\,(q\le p)$矩阵, 若$Y\perp\!\!\!\perp\{\mathsf{E}(Y^1|X),\cdots,\mathsf{E}(Y^j|X)\}|B^TX$, 则$\mathcal{S}_B$是一个$Y$关于$X$的前$j$阶条件矩DRS. 令$\mathcal{S}^{(j)}_{Y|X}$是所有前$j$阶条件矩DRSs的交集, 若$\mathcal{S}^{(j)}_{Y|X}$也是前$j$阶条件矩DRS, 则称其为中心前$j$阶条件矩子空间. 假设$\widehat{\beta}_{k\widehat{m}_k}\,(k=1,2,\cdots,j)$为利用KMWPLS方法得到高维相依预测变量中心$k\,(k=1,2,\cdots,j)$阶条件矩子空间方向的估计, 记矩阵$\widehat{B}=(\widehat{\beta}_{1\widehat{m}_1},\cdots,\widehat{\beta}_{j\widehat{m}_j})$, 然后对矩阵$\widehat{B}$进行奇异值分解, 令$\sigma_1\ge\sigma_2\ge\cdots\ge\sigma_r>0$是矩阵$\widehat{B}$的正奇异值, $l_1,l_2,\cdots,l_r$为对应的左奇异向量. 若中心前j阶条件矩子空间的维数为d, 则前d个左奇异向量所张成的子空间就是中心前j阶条件矩子空间的估计, $l_i\,(i=1,2,\cdots,d)$就是该空间的第i个预测向量. 至于实际问题中维数d的估计和检验的思想和方法, 可详细参见Li (1991), Cook和Weisberg (1991)所研究的置换检验方法以及Zhu, Miao和Peng (2006)提出的BIC估计方法等.

有关在高维相依情形, 基于k阶矩加权偏最小二乘估计方法的渐近正态性及证明还有待进一步研究和讨论.

§6. 定理证明

引理3.1的证明可参看Yin和Cook (2002)命题2, 此略.

定理3.1的证明: 因为由条件3易得, $\Sigma^{-1}\beta_{k0}\subseteq\mathrm{span}\{\gamma_1,\cdots,\gamma_m\}$. 又因为$\beta_{k0},\Sigma\beta_{k0},\cdots,\Sigma^{m-1}\beta_{k0}$线性独立, 而且它们中每一个都属于子空间$\mathrm{span}\{\gamma_1,\cdots,\gamma_m\}$的元素. 所以我们可以得到当$u=m$时$\mathrm{span}\{R_u\}=\mathrm{span}\{\gamma_1,\cdots,\gamma_m\}$. 从而$\Sigma^{-1}\beta_{k0}\subseteq\mathrm{span}\{R_u\}$.

注意到记号$P_{R_u}(\Sigma)=R_u(R_u^T\Sigma R_u)^{-1}R_u^T\Sigma$表示在关于$\Sigma$内积下到空间$\mathrm{span}\{R_u\}$的投影算子, 再结合刚才得到的结论$\Sigma^{-1}\beta_{k0}\subseteq\mathrm{span}\{R_u\}$. 我们就有

$$\beta_{ku}=R_u(R_u^T\Sigma R_u)^{-1}R_u^T\beta_{k0}=P_{R_u}(\Sigma)\Sigma^{-1}\beta_{k0}=\Sigma^{-1}\beta_{k0}\subseteq\mathcal{S}_{\mathsf{E}(Y^k|X)}.$$

于是该定理得证. □

为了证明定理3.2估计的强相合性, 先引进如下引理:

引理 6.1 (1) 设$\max\limits_{1\le i\le p}\mathsf{E}(|X_i|^2)<\infty$关于$p$一致成立, 且$p$随着$n$的增加而增大, 则

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}[x_i-\mathsf{E}(X)]\Big\|=o(\sqrt{p/n}\log n),\qquad \text{几乎处处成立;}\tag{6.1}$$

(2) 设$\max\limits_{1\le i\le p}\mathsf{E}(|X_i|^4)<\infty$于$p$一致成立, 且$p$随着$n$的增加而增大, 则

$$\Big\|\frac{1}{n}\sum_{i=1}^{n}[x_ix_i^T-\mathsf{E}(XX^T)]\Big\|=o(p\log n/\sqrt{n}),\qquad \text{几乎处处成立.}\tag{6.2}$$

该引理证明可参看Zhu和Zhu (2009) Lemma 1类似证明. 此略.

下面给出定理3.2的证明.

定理3.2的证明: 我们首先讨论矩估计$\widehat{\Sigma}$和$\widehat{\beta}_{k0}$的收敛速度. 由于假设$\mathsf{E}(X)=0$, 故

$$\begin{aligned}\|\widehat{\Sigma}-\Sigma\| &= \Big\|\frac{1}{n}\sum_{i=1}^n x_ix_i^T-\mathsf{E}(XX^T)-\frac{1}{n}\sum_{i=1}^n x_i\frac{1}{n}\sum_{i=1}^n x_i^T\Big\| \\ &\le \Big\|\frac{1}{n}\sum_{i=1}^n x_ix_i^T-\mathsf{E}(XX^T)\Big\|+\Big\|\frac{1}{n}\sum_{i=1}^n x_i\frac{1}{n}\sum_{i=1}^n x_i^T\Big\| \\ &= S_1+S_2.\end{aligned}$$

而根据引理6.1可知$S_1=o(p\log n/\sqrt{n})$, $S_2=\big\|(1/n)\cdot\sum\limits_{i=1}^n x_i\big\|^2=o(p\log^2 n/n)$几乎处处成立. 因此, 我们得到

$$\|\widehat{\Sigma}-\Sigma\|=o(p\log n/\sqrt{n}),\qquad \text{几乎处处成立.} \tag{6.3}$$

由引理6.1类似易得

$$\|\widehat{\beta}_{k0}-\beta_{k0}\|=\Big\|\frac{1}{n}\sum_{i=1}^n[x_iy_i^k-\mathsf{E}(XY^k)]\Big\|=o(\sqrt{p/n}\log n),\qquad \text{几乎处处成立.} \tag{6.4}$$

接下来我们证明基于k阶矩加权偏最小二乘(KMWPLS)估计$\widehat{\beta}_{k\widehat{m}_k}$的强相合性, 其中$\widehat{\beta}_{k\widehat{m}_k}=\widehat{R}_{\widehat{m}_k}(\widehat{R}_{\widehat{m}_k}^T\widehat{\Sigma}\widehat{R}_{\widehat{m}_k})^{-1}\widehat{R}_{\widehat{m}_k}^T\widehat{\beta}_{k0}$.

由于

$$\begin{aligned}\mathsf{P}(\|\widehat{\beta}_{k\widehat{m}_k}-\beta_{km}\|\ge pn^{-1/2}\log n) &= \mathsf{P}(\|\widehat{\beta}_{k\widehat{m}_k}-\beta_{km}\|\ge pn^{-1/2}\log n,\widehat{m}_k=m) \\ &\quad +\mathsf{P}(\|\widehat{\beta}_{k\widehat{m}_k}-\beta_{km}\|\ge pn^{-1/2}\log n,\widehat{m}_k\ne m) \\ &\le \mathsf{P}(\|\widehat{\beta}_{km}-\beta_{km}\|\ge pn^{-1/2}\log n)+\mathsf{P}(\widehat{m}_k\ne m) \\ &=: I_1(n)+I_2(n).\end{aligned}$$

因此, 我们可以通过分别讨论$I_1(n)$和$I_2(n)$的收敛性来证明估计收敛, 即相合的. 由于$\widehat{m}_k\to m$意味着$\sum\limits_{n=1}^{\infty}I_2(n)<\infty$. 故接下来我们只需要证明$\sum\limits_{n=1}^{\infty}I_1(n)<\infty$, 也即证明$\|\widehat{\beta}_{km}-\beta_{km}\|=o(pn^{-1/2}\log n)$几乎处处成立.

注意到第一步证明中的结论(6.3)式和(6.4)式, 结合三角不等式和矩阵Frobenius范数的相容性条件, 我们可以得到

$$\begin{aligned}\|\widehat{\Sigma}^i\widehat{\beta}_{k0}-\Sigma^i\beta_{k0}\| &\le \|(\widehat{\Sigma}^i-\Sigma^i)\widehat{\beta}_{k0}\|+\|\Sigma^i(\widehat{\beta}_{k0}-\beta_{k0})\| \\ &\le \|\widehat{\Sigma}^i-\Sigma^i\|\|\widehat{\beta}_{k0}\|+\|\Sigma^i\|\|\widehat{\beta}_{k0}-\beta_{k0}\| \\ &= o(pn^{-1/2}\log n),\qquad \text{对 } 1\le i\le m \text{ 几乎处处成立.}\end{aligned}$$

从而根据$R_m=(\beta_{k0},\Sigma\beta_{k0},\cdots,\Sigma^{m-1}\beta_{k0})$的定义和对应的估计

$$\widehat{R}_m=(\widehat{\beta}_{k0},\widehat{\Sigma}\widehat{\beta}_{k0},\cdots,\widehat{\Sigma}^{m-1}\widehat{\beta}_{k0}).$$

并结合三角不等式和矩阵Frobenius范数的相容性条件等性质, 类似易证

$$\begin{aligned} &\|\widehat{R}_m - R_m\| = o(pn^{-1/2}\log n), \\ &\|\widehat{R}_m^T\widehat{\Sigma}\widehat{R}_m - R_m^T\Sigma R_m\| = o(pn^{-1/2}\log n). \end{aligned} \tag{6.5}$$

于是根据矩阵的性质, 对任意非奇异矩阵U和对应的估计$\widehat{U}$,

$$\widehat{U}^{-1} - U^{-1} = U^{-1}(U - \widehat{U})\widehat{U}^{-1}.$$

我们易得

$$\|(\widehat{R}_m^T\widehat{\Sigma}\widehat{R}_m)^{-1} - (R_m^T\Sigma R_m)^{-1}\| = o(pn^{-1/2}\log n), \qquad \text{几乎处处成立.} \tag{6.6}$$

再利用以上类似的证明方法可证$\|\widehat{\beta}_{km} - \beta_{km}\| = o(pn^{-1/2}\log n)$几乎处处成立. 因此由Borel-Cantelli引理即得$\sum\limits_{n=1}^{\infty} I_1(n) < \infty$.

综上可得相合性得证. □

参 考 文 献

[1] Cook, R.D., *Regression Graphics: Ideas for Studying Regressions through Graphics*, Wiley & Sons, New York, 1998.

[2] Cook, R.D. and Weisberg, S., Graphs in statistical analyses: is the medium the message? *Amer. Statist.*, **53**(1999), 29–37.

[3] Cook, R.D. and Li, B., Dimension reduction for conditional mean in regression, *Ann. Statist.*, **30** (2002), 455–474.

[4] Yin, X. and Cook, R.D., Dimension reduction for conditional k-th moment in regression, *J. Roy. Statist. Soc., Ser. B*, **64**(2002), 159–175.

[5] Zhu, L.P. and Zhu, L.X., On distribution-weighted partial least squares with diverging number of highly correlated predictors, *J. Roy. Statist. Soc., Ser. B*, **71**(2009), 525–548.

[6] Cook, R.D. and Lee, H., Dimension reduction in regressions with a binary response, *J. Amer. Statist. Assoc.*, **94**(1999), 1187–1200.

[7] Wold, H., Estimating of principal components and related models by iterative least squares, In *Multi Analysis*, Ed. Krishnaiah, P.R., 391–420, New York: Academic Press, 1966.

[8] Wold, H., Soft modelling by latend variables: the nonlinear partial least squares (NIPALS) approach, In *Perspectives in Probability and Statistics, Papers in Honor of M.S. Barlett*, Ed. Gani, J., 117–142, London: Academic Press, 1975.

[9] Wold, S., Martens, H. and Wold, H., The multivariate calibration problem solved by the PLS method, *Proc. Conf. Matrix Pencils*, Eds. Ruhe, A. and Kågström, B., 286–293, Lecture notes in mathematics, Springer Verlag, Heidelberg, 1983.

[10] Naik, P. and Tsai, C.L., Partial least squares estimator for single-index models, *J. Roy. Statist. Soc., Ser. B*, **62**(2000), 763–771.

[11] Helland, I.S., Partial least squares regression and statistical models, *Scand. J. Statist.*, **17**(1990), 97–114.

[12] Li, K.C., Sliced inverse regression for dimension reduction (with discussion), *J. Amer. Statist. Assoc.*, **86**(1991), 316–342.

[13] Cook, R.D. and Weisberg, S., Discussion to "Sliced inverse regression for dimension reduction", *J. Amer. Statist. Assoc.*, **86**(1991), 316–342.

[14] Zhu, L.X., Miao, B.Q. and Peng, H., On sliced inverse regression with high-dimensional covariates, *J. Amer. Statist. Assoc.*, **101**(2006), 630–643.

The Central kth-Conditional Moment Suspace Estimation with Highly Dimensional and Highly Correlated Predictors

Xu Qunfang

(*Department of Statistics, Zhejiang Agriculture and Forset University, Lin'an, 311300*)

The conditional mean, variance and higher-conditional moment functions are often of special interest in regression. In this paper,we generalize central mean subspace and focus especial attention on the kth-conditional moment function. For this, we first borrow the new concept — the central kth-conditional moment subspace, and study its basic properties. To avoid computing the inverse of the covariance of predictors with large dimensionality and highly collinearity, we develop a method called the kth-moment weighted partial least squares to handle with the estimation of the central kth-conditional moment subspace. Finally, we obtain strong consistency.

Keywords: Sufficient dimension reduction subspace, central kth-conditional moment subspace, high dimensionality and collinearity, least squares estimation, partial least squares.

AMS Subject Classification: 62G05.