DOI:10.13718/j.cnki.xsxb.2017.01.003

基于累积切片的多元响应降维子空间估计①

甘胜进[1], 涂开仁[2], 游文杰[1]

1. 福建师范大学福清分校 电子与信息工程学院,福建 福清 350300;
2. 福建师范大学福清分校 经济与管理学院,福建 福清 350300

摘要:将相关研究提出的累积切片均值估计(CUME)、累积海塞方向(CHD)应用到多元响应降维子空间的估计中并对估计量加以改进得到改进的多元累积切片估计(g-CUME)、多元累积海塞方向(g-CHD).

关 键 词:多元响应降维子空间;切片逆回归;累积切片估计;海塞主方向;累积海塞估计

中图分类号:O213 **文献标志码**:A **文章编号**:1000-5471(2017)01-0013-07

随着大数据时代的来临,几乎在每个科学领域都会产生大数据,大数据显著特点之一就是高维.在高维空间中进行统计建模,往往耗费大量的计算时间,而且模型容易出现较大的偏差,这就是所谓的"维数灾难"现象.本质上讲,这是由高维数造成的,如何对变量的维数作有效降维处理显得尤为重要.常见的降维方法有主成分回归分析、偏最小二乘回归、投影寻踪等.主成分回归分析仅仅考虑了自变量之间的相关信息,忽略了与因变量之间的关系,而偏最小二乘虽然同时考虑自变量与因变量之间的相关关系,但是仅仅局限于线性关系,没有考虑非线性关系,另外投影寻踪需要估计连接函数,超出数据预处理的范围,不在考虑之内.

充分降维是近年来降维领域兴起的一种新方法.首先其对模型不作特定假设,即无模型建模(model-free),仅仅对自变量边缘分布作一些限制性假设,而这些限制性假设在高维空间中通常是近似满足[1];其次降维幅度大,降维之后的变量一般用少数几个原变量的线性组合来替代.其基本思路如下:对于给定一维响应变量 Y 和 p 维解释变量 $\boldsymbol{X}=(X_1, X_2, \cdots, X_p)^{\mathrm{T}}$,如果存在 $p\times k(k\leqslant p)$ 矩阵 $\boldsymbol{\eta}$ 满足 $Y\perp\!\!\!\perp \boldsymbol{X}\mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}$,其中 $\perp\!\!\!\perp$ 表示统计独立,$Y\perp\!\!\!\perp \boldsymbol{X}\mid \boldsymbol{U}$ 表示在任意给定 U 值的情况下,$\boldsymbol{X}$ 与 Y 相互独立.则有 $F_{Y|\boldsymbol{X}}(y\mid \boldsymbol{x})=F_{Y|\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}}(y\mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{x})$,这样 $\boldsymbol{X}$ 关于 Y 的回归信息就包含在 $\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}$ 当中,Y 对 $\boldsymbol{X}$ 条件分布依赖于 k 个自变量 $\boldsymbol{X}$ 的线性组合:$\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}$,$Y$ 对 $\boldsymbol{X}$ 条件分布维数不再是 p 维,而是 k 维,如果 k 远远小于 p,那么降维目的就达到了.特别当 $k=1$ 或 2 时,便可在概要图(summary plot)上分析 Y 与 $\boldsymbol{X}$ 之间的回归关系.由于 $Y\perp\!\!\!\perp \boldsymbol{X}\mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}\Leftrightarrow Y\perp\!\!\!\perp \boldsymbol{X}\mid (\boldsymbol{\eta}\boldsymbol{B})^{\mathrm{T}}\boldsymbol{X}$,其中 $\boldsymbol{B}$ 为 k 阶可逆方阵,$\boldsymbol{\eta}$ 与 $\boldsymbol{\eta}\boldsymbol{B}$ 张成子空间是相同的,所以我们关心的是 $span\{\boldsymbol{\eta}\}$,而不是 $\boldsymbol{\eta}$ 本身,$span\{\boldsymbol{\eta}\}$ 表示由 $\boldsymbol{\eta}$ 列向量张成子空间.借鉴充分统计量概念,称 $span\{\boldsymbol{\eta}\}$ 为一充分降维子空间(dimension reduction subspace, DRS).如果 $\bigcap \boldsymbol{\eta}$ 仍然是一充分降维子空间,则称之为中心降维子空间(central dimension reduction subspace, CS),记为 $S_{Y|\boldsymbol{X}}$,$rank(S_{Y|\boldsymbol{X}})$ 称为结构维数,显然 CS 是结构维数最小的降维子空间,有时候感兴趣的是 $E(Y\mid \boldsymbol{X})$,那么 $Y\mid \boldsymbol{X}$ 条件分布变得冗余,Cook 和 Li[2] 提出均值降维子空间,即:

$$Y\perp\!\!\!\perp E(Y\mid \boldsymbol{X})\mid \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} \tag{1}$$

类似 CS 讨论,若所有满足(1)式的集合的交集仍然满足(1)式,则称其为中心均值降维子空间(central mean dimension reduction subspace, CMS),记为 $S_{E(Y|X)}$.令 $\boldsymbol{Z}=\boldsymbol{\Sigma}^{-\frac{1}{2}}(\boldsymbol{X}-E(\boldsymbol{X}))$,则有转换公式 $S_{Y|\boldsymbol{X}}=$

① 收稿日期:2016-04-26
基金项目:国家自然科学基金面上项目(61473329);福建省自然科学基金面上项目(2015J01009);福建省中青年教师教育科研项目(JAT160566).
作者简介:甘胜进(1982-),男,湖北黄冈人,硕士,讲师,主要从事高维数据分析研究.

$\boldsymbol{\Sigma}^{-\frac{1}{2}}S_{Y|\boldsymbol{Z}}$[3]，其中 $\boldsymbol{\Sigma}=D(\boldsymbol{X})>0$. 本文假定 $E(\boldsymbol{X})=0$，$D(\boldsymbol{X})=\boldsymbol{I}_p$.

通常估计降维子空间所需的两个基本条件为：

1) 线性条件均值：$E(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})$ 为 $\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}$ 线性函数，即 $E(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})=\boldsymbol{P}_{\boldsymbol{\eta}}\boldsymbol{X}$，$\forall\,\boldsymbol{\eta}\in\boldsymbol{R}^p$，其中投影阵 $\boldsymbol{P}_{\boldsymbol{\eta}}=\boldsymbol{\eta}(\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^{\mathrm{T}}$.

2) 常数条件方差：$Var(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})$ 为非随机矩阵，即 $Var(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})=\boldsymbol{I}-\boldsymbol{P}_{\boldsymbol{\eta}}$.

1 累积切片估计相关理论

类似一元响应变量讨论，如果把它推广到多维响应变量，那么就是多元响应降维子空间、多元响应均值降维子空间，即：

$$\boldsymbol{Y}\perp\!\!\!\perp\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} \tag{2}$$

$$\boldsymbol{Y}\perp\!\!\!\perp E(\boldsymbol{Y}\mid\boldsymbol{X})\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X} \tag{3}$$

其中：$\boldsymbol{X}=(X_1,X_2,\cdots,X_p)^{\mathrm{T}}$，$\boldsymbol{Y}=(Y_1,Y_2,\cdots,Y_q)^{\mathrm{T}}$.

在线性条件均值下，$E(\boldsymbol{X}\mid Y)\subseteq span\{\boldsymbol{\eta}\}$. 为估计 $E(\boldsymbol{X}\mid Y)$，利用经典的切片逆回归方法(slicedInverse regression，SIR)[4] 对响应变量 Y 进行切片：设 $\{(\boldsymbol{X}_i,Y_i)\}_{i=1}^n$ 为总体 $(\boldsymbol{X},Y)$ 的样本，将 Y 的取值范围划分成 h 个区间(切片)：$I_1,I_2,\cdots,I_h$，$I_i=(a_{i-1},a_i]$，$-\infty=a_0<a_1<a_2\cdots<a_h=+\infty$，计算各个切片上 $\boldsymbol{X}$ 的样本均值 $E_n(\boldsymbol{X}\mid Y\in I_i)=\dfrac{E_n(\boldsymbol{X}I(Y\in I_i))}{E_n(I(Y\in I_i))}$，其中 $E_n(f(x))=\dfrac{1}{n}\sum\limits_{j=1}^{n}f(x_j)$，$I(\cdot)$ 表示示性函数. 如果结构维数 d 已知，取核矩阵 $\boldsymbol{M}=\sum\limits_{i=1}^{h}E_n(\boldsymbol{X}\mid Y\in I_i)(E_n(\boldsymbol{X}\mid Y\in I_i))^{\mathrm{T}}E_n(I(Y\in I_i))$ 前 d 个最大特征值对应的特征向量为中心降维子空间的估计. 实际上切片估计量 $E_n(\boldsymbol{X}\mid Y\in I_i)$ 是用来估计 $E(\boldsymbol{X}\mid Y\in I_i)$，而非 $E(\boldsymbol{X}\mid Y)$，但是实际当中，只要切片数量选取合理，用 $E_n(\boldsymbol{X}\mid Y\in I_i)$ 近似 $E(\boldsymbol{X}\mid Y)$ 的效果非常好. 特别当切片 $h=2$ 时，$\boldsymbol{M}=\dfrac{E_n(\boldsymbol{X}I(Y\in I_1))(E_n(\boldsymbol{X}I(Y\in I_1)))^{\mathrm{T}}}{E_n(I(Y\in I_1))-(E_n(I(Y\in I_1)))^2}$，只能得到一个估计方向，Zhu[5] 从中得到启发，构造了估计量 $\boldsymbol{m}(y)=E(\boldsymbol{X}I(Y\in I_1))$，其中 $I_1=(-\infty,y]$，在线性条件均值下，$\boldsymbol{m}(y)\subseteq span\{\boldsymbol{\eta}\}$，让 y 取遍 Y 的所有可能值，然后求平均，于是得到基于二切片的总体中心降维子空间估计量 $\widetilde{\boldsymbol{M}}=E(\boldsymbol{m}(Y)\boldsymbol{m}(Y)^{\mathrm{T}}w(Y))$，其中 $w(Y)$ 为权重函数，通过统计模拟发现，估计效果与权重函数无关，因此为计算方便，舍去 $w(Y)$ 采用 $\widetilde{\boldsymbol{M}}=E(\boldsymbol{m}(Y)\boldsymbol{m}(Y)^{\mathrm{T}})$，称该方法为累积均值估计(cumulative mean estimation，CUME)，通过大量的模拟研究发现，相比于SIR，SAVE(sliced average variance estimation，切片平均方差估计)[6]，DR(directional regression，方向回归)[7]，在迹相关准则下[8]，CUME 估计的方向更加靠近真实方向，而且比较稳定. 本文把一维的 CUME 方法推广到多维累积均值估计(简记 m-CUME)，针对 m-CUME 对对称回归函数失效问题，推广一维的累积海塞方向(cumulative Hessian directions，CHD)[9] 到多维的累积海塞方向(m-CHD)，不过在构造样本估计量方面，采用 Feng[10] 提出的方法. 下面讨论 m-CUME 和 m-CHD 相关性质以及样本估计量大样本性质.

定理 1 在线性条件均值下，m-CUME：$\boldsymbol{M}_{m-CUME}=E(\boldsymbol{m}(\boldsymbol{Y})\boldsymbol{m}(\boldsymbol{Y})^{\mathrm{T}})\subseteq span\{\boldsymbol{\eta}\}$.

证 $\boldsymbol{m}(\boldsymbol{y})=E(\boldsymbol{X}\boldsymbol{I}(\boldsymbol{Y}\leqslant\boldsymbol{y}))=E(E(\boldsymbol{X}\mid\boldsymbol{Y})I(Y\leqslant\boldsymbol{y}))$

而

$$E(\boldsymbol{X}\mid\boldsymbol{Y})=E(E(\boldsymbol{X}\mid\boldsymbol{Y},\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})\mid\boldsymbol{Y})=E(E(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})\mid\boldsymbol{Y})=\boldsymbol{P}_{\boldsymbol{\eta}}E(\boldsymbol{X}\mid\boldsymbol{Y}) \tag{4}$$

其中：$\boldsymbol{y}=(y_1,y_2,\cdots,y_q)\in\mathbf{R}^q$，(4)式第一个等式用到了条件概率平滑性，第二个等式用到了(2)式，最后一个等式用到了线性条件均值条件. 故

$$\boldsymbol{m}(\boldsymbol{y})=\boldsymbol{P}_{\boldsymbol{\eta}}E(E(\boldsymbol{X}\mid\boldsymbol{Y})I(\boldsymbol{Y}\leqslant\boldsymbol{y}))=\boldsymbol{P}_{\boldsymbol{\eta}}E(\boldsymbol{X}\boldsymbol{I}(\boldsymbol{Y}\leqslant\boldsymbol{y}))\subseteq span\{\boldsymbol{\eta}\}$$

证毕.

m-CUME 样本估计：样本 $(\boldsymbol{X}_i,\boldsymbol{Y}_i)_{i=1}^n$，其中

$$\boldsymbol{X}_i=(X_{i1},X_{i2},\cdots,X_{ip})^{\mathrm{T}},\ \boldsymbol{Y}_i=(Y_{i1},Y_{i2},\cdots,Y_{iq})^{\mathrm{T}}$$

$\widehat{\boldsymbol{m}}(\boldsymbol{y})=E_n(\boldsymbol{X}\boldsymbol{I}(\boldsymbol{Y}\leqslant\boldsymbol{y}))=\dfrac{1}{n}\sum\limits_{i=1}^{n}\boldsymbol{X}_iI(\boldsymbol{Y}_i\leqslant\boldsymbol{y})$，$\widehat{\boldsymbol{M}}_{m-CUME}=\dfrac{1}{n}\sum\limits_{i=1}^{n}\widehat{\boldsymbol{m}}(\boldsymbol{Y}_i)\widehat{m}(\boldsymbol{Y}_i)^{\mathrm{T}}$，类似文献[5]中方法得到该

估计量的强相合性质(定理 2)：

定理 2　假设 $\max\limits_{1\leqslant i\leqslant p}\mid X_i\mid^p<+\infty$ 关于自变量维数 p 一致成立，则

$$\|\widehat{\boldsymbol{M}}_{m-CUME}-\boldsymbol{M}_{m-CUME}\|=o\left(\frac{p}{\sqrt{n}}\ln n\right)$$

几乎必然成立，其中 $\|\cdot\|$ 表示 Frobenius 范数.

定理 2 表明 $\widehat{\boldsymbol{M}}_{m-CUME}$ 收敛 $\boldsymbol{M}_{m-CUME}$ 的速度依赖于自变量维数 p 和样本容量 n，p 越大收敛速度就越慢，而 n 越大收敛速度就越快. 由于在估计 $\widehat{\boldsymbol{m}}(\boldsymbol{Y}_i)$ 时，有可能大量的 $E_n(I(\boldsymbol{Y}\leqslant\boldsymbol{Y}_i))$ 很小，因此导致估计的 $\widehat{\boldsymbol{m}}(\boldsymbol{Y}_i)$ 不稳定，为了保证估计 $\widehat{\boldsymbol{m}}(\boldsymbol{Y}_i)$ 时拥有一定数量 $\boldsymbol{X}$ 的样本，采用文献[10]中方法得到 $\boldsymbol{M}_{g-CUME}$，即用

$$\widehat{\boldsymbol{m}}(Y_{ik})=\frac{1}{n}\sum_{j=1}^{n}\boldsymbol{X}_jI(Y_{jk}\leqslant Y_{ik})\qquad\widehat{\boldsymbol{M}}_{g-CUME}=\frac{1}{n}\sum_{k=1}^{q}\sum_{i=1}^{n}\widehat{\boldsymbol{m}}(Y_{ik})\widehat{\boldsymbol{m}}(Y_{ik})^{\mathrm{T}}$$

来估计 $\boldsymbol{M}_{m-CUME}$，尽管不是 $\boldsymbol{M}_{m-CUME}$ 的无偏估计，但是 $E(\widehat{\boldsymbol{M}}_{g-CUME})\subseteq span\{\boldsymbol{\eta}\}$，并且模拟结果表明其估计性能有大幅度提高.

当 $\boldsymbol{Y}$ 对 $\boldsymbol{X}$ 回归函数是对称函数时，$E(\boldsymbol{X}\mid\boldsymbol{Y})=0$，此时 m-CUME 估计往往失效. 针对这类问题，Yin 和 Bura[11] 提出 $E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\otimes(\boldsymbol{Y}-E(\boldsymbol{Y}))^{\mathrm{T}})$(简记为 YB 法)来估计多维中心降维子空间，Li[12] 提出投影抽样(projective resampling，PR)方法，即把多维因变量投影到一维空间上，然后在一维空间上采用 SIR，SAVE 等，由此产生的方法简记为 PRSIR，PRSAVE 等. Zhang[9] 将累积切片估计的思想应用到海塞主方向(principal hessian directions，PHD)[13] 上，提出累积的海塞方向(cumulative hessian directions，CHD)方法，与 PHD，SIR，SAVE，DR 相比，CHD 处理对称的回归函数更加有效. 其实 CHD 对于多维响应同样适用，简记为 m-CHD，其方法如下：设 $F(y)$ 为 $\boldsymbol{Y}$ 的边缘分布函数，

$$\boldsymbol{M}(\boldsymbol{y})=E((I(\boldsymbol{Y}\leqslant\boldsymbol{y})-F(\boldsymbol{y}))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}),\ \boldsymbol{M}_{m-CHD}=E(\boldsymbol{M}(\boldsymbol{Y})\boldsymbol{M}(\boldsymbol{Y})^{\mathrm{T}})$$

定理 3　在线性条件均值和常数条件方差下 $\boldsymbol{M}_{m-CHD}=E(\boldsymbol{M}(\boldsymbol{Y})\boldsymbol{M}(\boldsymbol{Y})^{\mathrm{T}})\subseteq span\{\boldsymbol{\eta}\}$.

证　$$\begin{aligned}\boldsymbol{M}(\boldsymbol{y})=&E((I(\boldsymbol{Y}\leqslant\boldsymbol{y})-F(\boldsymbol{y}))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=E((E(I(\boldsymbol{Y}\leqslant\boldsymbol{y})\mid\boldsymbol{X})-F(\boldsymbol{y}))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=\\&E((E(I(\boldsymbol{Y}\leqslant\boldsymbol{y})\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})-F(\boldsymbol{y}))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=E((I(\boldsymbol{Y}\leqslant\boldsymbol{y})-F(\boldsymbol{y}))E(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}))=\\&E((I(\boldsymbol{Y}\leqslant\boldsymbol{y})-F(\boldsymbol{y}))(Var(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})+E(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X})(E(\boldsymbol{X}\mid\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{X}))^{\mathrm{T}}))=\\&\boldsymbol{P}_{\boldsymbol{\eta}}E((I(\boldsymbol{Y}\leqslant\boldsymbol{y})-F(\boldsymbol{y}))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{P}_{\boldsymbol{\eta}}\subseteq span\{\boldsymbol{\eta}\},\text{ 故 }\boldsymbol{M}_{m-CHD}\subseteq span\{\boldsymbol{\eta}\}\end{aligned}$$

$\boldsymbol{M}_{m-CHD}$ 样本估计为：

$$\widehat{\boldsymbol{M}}(y)=\frac{1}{n}\sum_{i=1}^{n}(I(\boldsymbol{Y}_i\leqslant\boldsymbol{y})-F_n(\boldsymbol{y}))\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}},\ \widehat{\boldsymbol{M}}_{m-CHD}=\frac{1}{n}\sum_{i=1}^{n}\widehat{\boldsymbol{M}}(\boldsymbol{Y}_i)\widehat{\boldsymbol{M}}(\boldsymbol{Y}_i)^{\mathrm{T}}$$

其中：$F_n(\boldsymbol{y})$ 是 $\boldsymbol{Y}$ 的经验分布函数. 估计量 $\widehat{\boldsymbol{M}}_{m-CHD}$ 的强相合性质与 CHD 一样，详细证明参看文献[9]. 实际估计时面临与 $\widehat{\boldsymbol{M}}_{m-CUME}$ 同样的问题，大量的 $I(\boldsymbol{Y}_i\leqslant\boldsymbol{y})$ 与 $F_n(\boldsymbol{y})$ 可能都很小，甚至为 0，因此导致估计出现较大偏差和不稳定，采用类似改进 $\widehat{\boldsymbol{M}}_{m-CUME}$ 的方法得到改进的估计量：

$$\widehat{\boldsymbol{M}}(y_k)=\frac{1}{n}\sum_{i=1}^{n}(I(Y_{ik}\leqslant y_k)-F_n(y_k))\boldsymbol{X}_i\boldsymbol{X}_i^{\mathrm{T}},\ \widehat{\boldsymbol{M}}_{g-CHD}=\frac{1}{n}\sum_{k=1}^{q}\sum_{i=1}^{n}\widehat{\boldsymbol{M}}(Y_{ik})\widehat{\boldsymbol{M}}(Y_{ik})^{\mathrm{T}}$$

其中 $y_k\in R$，$F_n(y_k)$ 是 $\boldsymbol{Y}$ 的第 k 个分量的经验分布函数. 由于在实际当中，多元经验分布函数的计算比较耗时，由一维的 PHD 核矩阵可得其另外一种等价形式：

$$E((Y-E(Y))\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})=E(Y\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}})-E(YI)=E(Y(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{I}))$$

因此尝试 $E(I(\boldsymbol{Y}\leqslant\boldsymbol{y})(\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{I}))$ 来替代 m-CHD 方法中的 $\boldsymbol{M}(\boldsymbol{y})$，或者借鉴 YB 法写成 $E((\boldsymbol{X}\boldsymbol{X}^{\mathrm{T}}-\boldsymbol{I})\otimes(\boldsymbol{Y}I(\boldsymbol{Y}\leqslant\boldsymbol{y}))^{\mathrm{T}})$，样本估计时也采用 Feng 的方法，但是模拟结果表明其估计效果还不如 m-CHD，g-CHD.

2　模拟研究

衡量估计方向与真实方向接近程度大小的文献较多，采用文献[14]中的准则，即 $d(\hat{\boldsymbol{\eta}},\boldsymbol{\eta})=\|\boldsymbol{P}_{\hat{\boldsymbol{\eta}}}-\boldsymbol{P}_{\boldsymbol{\eta}}\|$，其中 $\boldsymbol{P}_{\hat{\boldsymbol{\eta}}}$，$P_{\boldsymbol{\eta}}$ 分别表示估计方向 $\hat{\boldsymbol{\eta}}$ 和真实方向 $\boldsymbol{\eta}$ 的列向量构成的投影阵，d 为真实结构维数，$\|\cdot\|$ 表示欧几里得范数，即矩阵的最大奇异值，显然 $d(\hat{\boldsymbol{\eta}},\boldsymbol{\eta})$ 越小，$\hat{\boldsymbol{\eta}}$ 就越接近 $\boldsymbol{\eta}$，特别当 $d(\hat{\boldsymbol{\eta}},\boldsymbol{\eta})=0$ 时，

$span(\hat{\boldsymbol{\eta}}) = span(\boldsymbol{\eta})$. 为简便记，若无其它说明，以下例子的随机模拟均为100次.

例1 $Y_1 = \boldsymbol{\eta}_1^{\mathrm{T}}\boldsymbol{X} + \varepsilon_1$, $Y_2 = \boldsymbol{\eta}_2^{\mathrm{T}}\boldsymbol{X} + \varepsilon_2$, $Y_3 = \varepsilon_3$, $Y_4 = \varepsilon_4$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)^{\mathrm{T}} \sim N_4(0, \boldsymbol{\Sigma})$

$\boldsymbol{\eta}_1 = (1, 0, \cdots, 0)^{\mathrm{T}}$, $\boldsymbol{\eta}_2 = (0, 1, 0, \cdots, 0)^{\mathrm{T}}$, $\boldsymbol{\Sigma} = (\sigma_{ij})$, $\sigma_{ij} = \begin{cases} 1 & i = j \\ 0.5 & i \neq j \end{cases}$.

表1 基于例1在不同样本容量n和维数p下m-CUME, g-CUME的估计效果对比

方法	p	n			
		200	400	600	800
m-CUME	8	0.435 6(0.106 5)	0.317 0(0.072 4)	0.252 1(0.061 2)	0.225 5(0.048 7)
	15	0.584 4(0.097 0)	0.434 9(0.070 2)	0.356 2(0.063 0)	0.312 3(0.053 1)
	20	0.669 5(0.103 1)	0.504 8(0.068 4)	0.425 0(0.067 9)	0.365 1(0.055 1)
	25	0.690 2(0.077 5)	0.555 2(0.073 8)	0.451 4(0.066 1)	0.411 0(0.054 8)
g-CUME	8	0.307 5(0.074 5)	0.218 1(0.049 5)	0.173 7(0.039 2)	0.156 5(0.033 8)
	15	0.427 6(0.072 1)	0.307 0((0.048 4)	0.251 3(0.042 0)	0.218 7(0.037 9)
	20	0.498 3(0.083 3)	0.362 1(0.052 8)	0.290 7(0.048 4)	0.252 7(0.041 2)
	25	0.531 8(0.070 7)	0.401 3(0.059 0)	0.316 5(0.044 9)	0.281 6(0.037 7)

表2 基于例1在维数$p = 10$下多种方法估计效果对比

方 法	n			
	300	500	800	1 200
MS($h = 16$)	0.284 4(0.055 4)	0.215 5(0.043 6)	0.166 8(0.029 6)	0.131 6(0.027 2)
MS($h = 81$)	0.289 0(0.063 2)	0.206 3(0.042 8)	0.148 5(0.031 1)	0.109 7(0.022 0)
MS($h = 256$)	0.828 8(0.141 9)	0.346 7(0.085 0)	0.168 9(0.034 8)	0.129 1(0.025 3)
PRSIR($h = 4$)	0.939 6(0.066 5)	0.933 0(0.081 3)	0.931 0(0.092 6)	0.946 8(0.066 7)
PRSIR($h = 8$)	0.937 9(0.078 9)	0.944 9(0.061 9)	0.944 9(0.072 5)	0.946 4(0.076 0)
PRSIR($h = 16$)	0.949 5(0.061 4)	0.952 4(0.063 4)	0.933 6(0.086 1)	0.945 2(0.078 3)
PRSAVE($h = 4$)	0.946 8(0.067 4)	0.960 3(0.052 7)	0.951 4(0.073 2)	0.943 3(0.075 9)
PRSAVE($h = 8$)	0.967 5(0.056 7)	0.961 5(0.056 7)	0.954 6(0.071 9)	0.940 9(0.075 4)
PRSAVE($h = 16$)	0.979 8(0.030 6)	0.962 3(0.048 0)	0.953 3(0.074 5)	0.945 7(0.081 5)
g-CUME	0.280 8(0.049 8)	0.215 6(0.044 8)	0.176 6(0.036 9)	0.142 6(0.024 2)

表1为100次重复下，m-CUME, g-CUME估计方向与真实方向接近程度的均值(括号外数据)和标准差(括号内数据). 从表1中可以看出：当样本容量n固定、维数p变大时，估计方向与真实方向的距离越来越大，当维数p不变、样本容量n变大时，估计的效果越来越好，充分印证了定理2；在相同样本容量和维数下，g-CUME估计的均值和标准差总比m-CUME要小，说明g-CUME比m-CUME有效. 表2为几种方法与g-CUME的比较实验结果，其中：MS将响应变量的每个分量分别切成2片、3片、4片，故切片总数量分别为2^4,3^4,4^4；PRSIR和PRSAVE的切片数量分别选取为4,8,16. 从表2可知：PRSIR和PRSAVE表现最差，估计效果远不及MS,g-CUME，当样本容量较大时，MS表现很好，但是当样本容量较小时，MS估计效果易受切片数量选择的影响，而如何选择切片数量是一个公开的难题. 相比之下，g-CUME不用选择切片数量，表现比较稳定.

例2 $Y_1 = X_1 + \dfrac{3X_2}{0.5 + X_1^2} + \varepsilon_1$, $Y_2 = X_2 + \mathrm{e}^{0.5X_2} + \varepsilon_2$, $Y_3 = X_1 + X_2 + \varepsilon_3$, $Y_4 = \varepsilon_4$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)^{\mathrm{T}}$ 的分布同例1.

图1、图2是样本容量均为500，维数分别为20和40时，g-CUME与选取的几种方法在例2上的估计表现. 从图中可看出：PRSAVE表现最糟糕，其次是PRSIR. 当切片数量选取合理时，MS表现也不错，MS估计效果受切片数量的影响非常明显，相比之下，g-CUME是几种估计方法中表现最好的.

例3 $Y_1 = X_1^2 + \varepsilon_1$, $Y_2 = X_2\varepsilon_2$, $Y_3 = \varepsilon_3$, $Y_4 = \varepsilon_4$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4)^{\mathrm{T}}$ 的分布同例1.

例3中响应变量对解释变量的回归函数既有对称型，也有异方差型，从100次模拟结果(表3)中看出，g-CUME与m-CUME相比估计效果有较为明显的改善，从表4看出：MS和PRSIR表现最差，这在意料之

中，因为它们对对称型回归函数失效；当样本量较大时，PRSAVE 比 PRPHD 要略好；而 YB 虽然对对称型函数有效，但是对本例中异方差型无效，因此表现虽然略强于其它几种方法，但是与 g-CHD 相比，差距非常大，尤其是当样本容量增大时，这种差距逐渐拉大，g-CHD 方法的优势愈发明显.

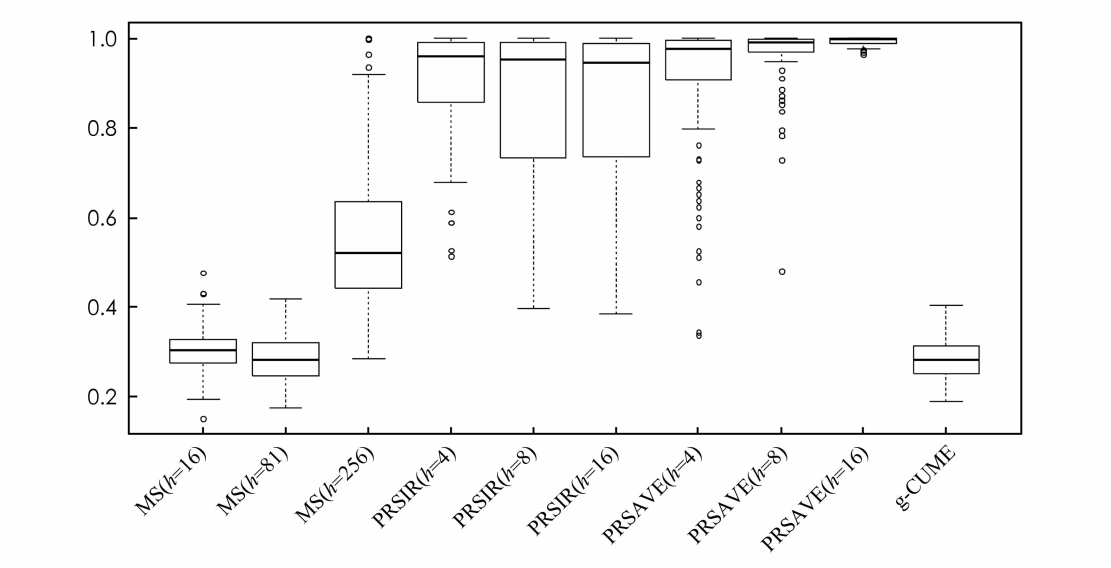

图 1　样本容量 $n = 500$ 和维数 $p = 20$，多种选取方法估计效果的箱线图

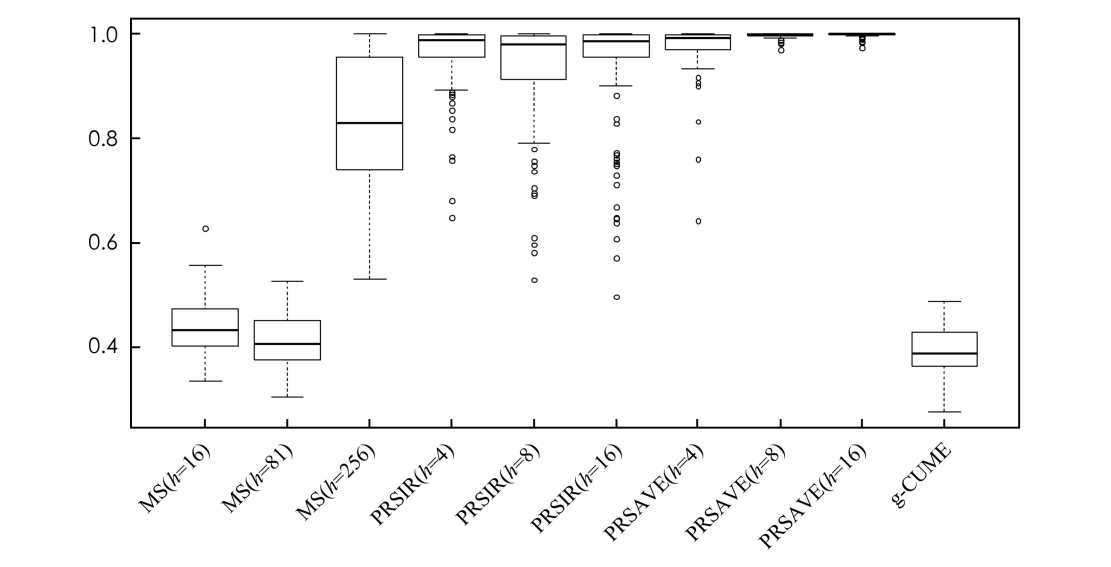

图 2　样本容量 $n = 500$ 和维数 $p = 40$，多种选取方法估计效果的箱线图

表 3　基于例 3 在不同样本容量 *n* 和维数 *p* 下 m-CUME，g-CUME 估计效果对比

方法	p	n			
		200	300	400	500
m-CUME	10	0.898 4(0.114 6)	0.839 6(0.148 6)	0.794 9(0.183 2)	0.750 2(0.206 3)
	15	0.964 2(0.046 7)	0.932 5(0.080 2)	0.912 8(0.089 6)	0.875 4(0.128 0)
	20	0.980 5(0.027 2)	0.969 4(0.038 6)	0.959 1(0.053 3)	0.936 4(0.069 2)
	25	0.988 4(0.019 6)	0.984 7(0.024 3)	0.972 2(0.036 1)	0.967 7(0.041 9)
g-CUME	10	0.795 7(0.168 0)	0.701 9(0.181 9)	0.607 2(0.182 6)	0.535 0(0.193 4)
	15	0.897 1(0.108 9)	0.837 1(0.144 7)	0.789 1(0.160 8)	0.711 9(0.165 8)
	20	0.936 2(0.069 8)	0.916 9(0.082 3)	0.861 5(0.124 3)	0.811 7(0.134 4)
	25	0.953 0(0.056 6)	0.938 9(0.067 3)	0.908 9(0.096 9)	0.887 0(0.104 5)

表4 基于例3在维数 $p=10$ 下多种方法与g-CUME估计效果比较

Method	n			
	200	400	600	800
MS($h=16$)	0.945 9(0.073 0)	0.944 3(0.067 2)	0.944 3(0.062 7)	0.943 6(0.077 6)
MS($h=81$)	0.935 7(0.095 4)	0.931 4(0.097 4)	0.921 8(0.105 8)	0.933 8(0.104 2)
MS($h=256$)	0.964 7(0.051 3)	0.919 8(0.100 0)	0.945 5(0.087 2)	0.922 6(0.099 7)
PRSIR($h=4$)	0.967 6(0.040 8)	0.966 5(0.045 4)	0.956 0(0.061 5)	0.964 7(0.057 6)
PRSIR($h=8$)	0.965 5(0.049 7)	0.966 0(0.049 2)	0.959 6(0.052 5)	0.963 3(0.051 9)
PRSIR($h=16$)	0.964 5(0.054 7)	0.957 4(0.058 7)	0.959 1(0.053 4)	0.956 4(0.057 4)
PRSAVE($h=4$)	0.923 4(0.119 3)	0.902 9(0.159 1)	0.850 3(0.201 1)	0.851 0(0.188 2)
PRSAVE($h=8$)	0.921 4(0.110 9)	0.896 0(0.152 8)	0.839 4(0.208 6)	0.852 3(0.201 1)
PRSAVE($h=16$)	0.904 9(0.107 9)	0.913 4(0.140 6)	0.869 8(0.175 9)	0.854 8(0.192 7)
PRPHD	0.9213(0.086 2)	0.929 0(0.077 0)	0.904 0(0.128 9)	0.896 9(0.123 4)
YB	0.880 4(0.118 4)	0.884 2(0.113 7)	0.961 2(0.154 6)	0.842 5(0.147 2)
g-CHD	0.817 5(0.144 6)	0.631 5(0.205 1)	0.434 6(0.155 7)	0.341 6(0.128 5)

例4 $Y_1=X_1+X_1^2+\varepsilon_1$，$Y_2=\cos(X_2)+\varepsilon_2$，$Y_3=\varepsilon_3$，$Y_4=\varepsilon_4$，$\boldsymbol{\varepsilon}=(\varepsilon_1,\varepsilon_2,\varepsilon_3,\varepsilon_4)^{\mathrm{T}}$ 的分布同例1.

例4中的回归函数既有线性又有对称性，从图3、图4看出：PRPHD，YB，g-CHD表现明显优于选取的其它几种方法，而三者中g-CHD表现又是最好的.

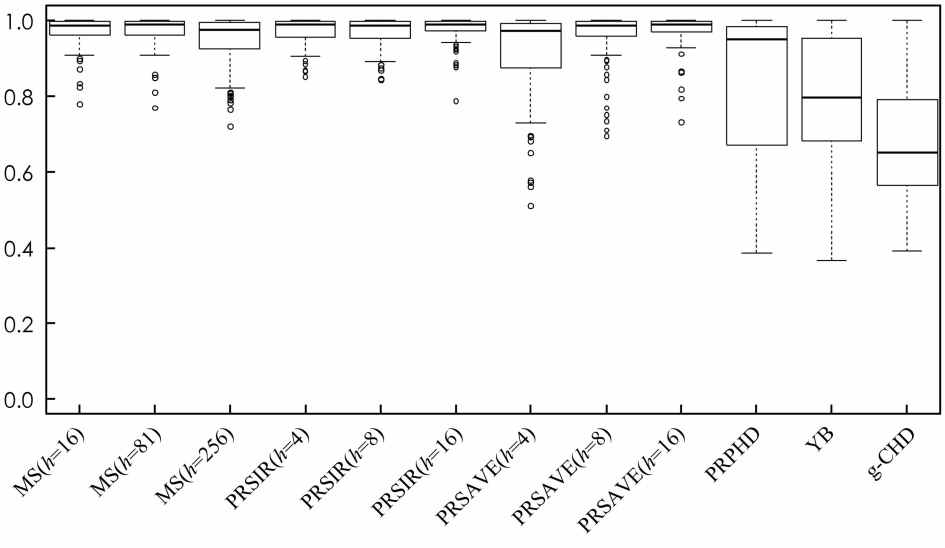

图3 样本容量 $n=500$ 和维数 $p=20$，多种选取方法估计效果的箱线图

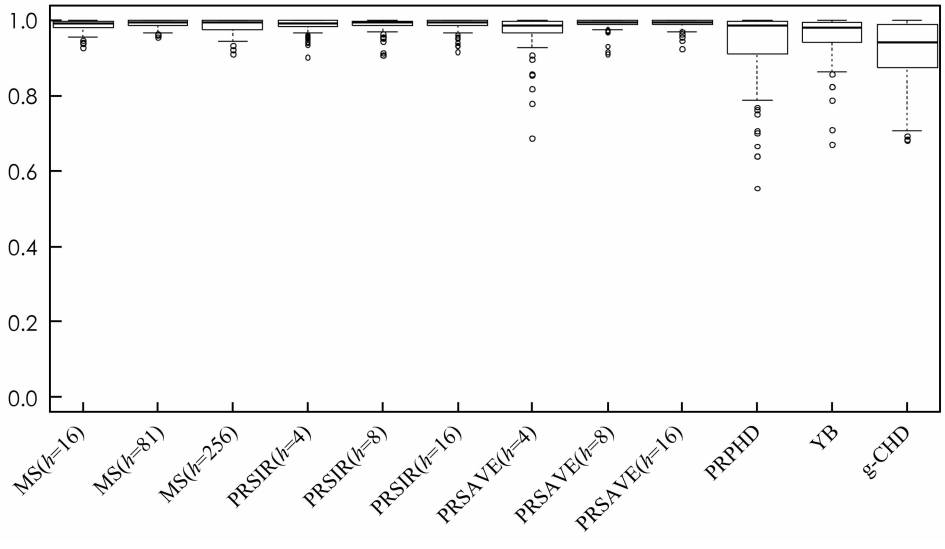

图4 样本容量 $n=500$ 和维数 $p=40$，多种选取方法估计效果的箱线图

3 结　　语

本文将一元响应的累积切片、海塞方向法应用到多元响应，并在估计量的构造方面适当加以改进，显著提高了估计效果. 然而当响应变量维数较大时，这种计算比较耗时，这是因为海塞主方向计算比较费时.

参考文献：

[1] HALL P, LI K C. On Almost Linearity of Low-Dimensional Projections from High-Dimensional Data [J]. Ann Statist, 1993, 21(2): 867－889.

[2] COOK R D, LI B. Dimension Reduction for Conditional Mean in Regression [J]. Ann Statist, 2002, 30(2): 455－474.

[3] COOK R D. Regression Graphics: Ideas for Studying Regressions Through Graphics [M]. New York: Wiley, 1998.

[4] LI K C. Sliced Inverse Regression for Dimension Reduction [J]. J Amer Statist Assoc, 1991, 86(414): 316－327.

[5] ZHU L P, ZHU L X, FENG Z H. Dimension Reduction in Regressions Through Cumulative Slicing Estimation [J]. J Amer Statist Assoc, 2010, 105(492): 1455－1466.

[6] COOK R D. Comment on the "Sliced Inverse Regression for Dimension Reduction" [J]. J Amer Statist Assoc, 1991, 86(414): 328－332.

[7] LI B, WANG S l. On Directional Regression for Dimension Reduction [J]. J Amer Statist Assoc, 2009, 102(479): 997－1008.

[8] FERRÉ L. Determining the Dimension in Sliced Inverse Regressionand Related Methods [J]. J Amer Statist Assoc, 1998, 93(441): 132－140.

[9] ZHANG L M, ZHU L P, ZHU L X. Sufficient Dimension Reduction in Regressions Through Cumulative [J]. Stat Comput, 2011, 21(3): 325－334.

[10] FENG Z H, WEN X M, YU Z, et al. On Partial Sufficient Dimension Reduction with Applications to Partially Linear Multi-Index Models [J]. J Amer Statist Assoc, 2013, 108(501): 237－246.

[11] YIN X R, BURA. Moment-based Dimension Reduction for Multivariate Response Regression [J]. J Statist Plann Inference, 2006, 136(10): 3675－3688.

[12] LI B, WEN S Q, ZHU L X. On a Projective Resampling Method for Dimension Reduction with Multivariate Responses [J]. J Amer Statist Assoc, 2008, 103(483): 1177－1186.

[13] LI K C. On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma [J]. J Amer Statist Assoc, 1992, 87(420): 1025－1039.

[14] LI B, CHIAROMONTE F. Contour Regression: A General Approach to Dimension Reduction [J]. Ann Statist, 2005, 33(4): 1580－1616.

Estimation For Multivariate Responses Central Dimension Reduction Subspaces Based On Cumulative Slicing

GAN Sheng-jin[1], TU Kai-ren[2], YOU Wen-jie[1]

1. *School of Electronical and Information Engineering, Fuqing Branch of Fujian Normal University, Fuqing Fujian 350300, China;*

2. *School of Economics and Management, Fuqing Branch of Fujian Normal University, Fuqing Fujian 350300, China*

Abstract: Based on Cumulative Slicing Estimation(CUME) and Cumulative Hessian Directions(CHD) for univariate response, multivariate responses Cumulative Slicing Estimation(m-CUME) and multivariate responses Cumulative Hessian Directions(m-CHD) have been proposed to estimate multivariate responses central dimension reduction subspaces, modified m-CUME and m-CHD which are denoted by g-CUME and g-CHD respectively outperform theirs through statistical simulations, are also comparable with selected methods.

Key words: multivariate responses dimension reduction subspaces; sliced inversed regression; cumulative slicing estimation; principal Hessian directions; cumulative Hessian directions

责任编辑　张　枸