第45卷第9期 中国科学技术大学学报 Vol. 45, No. 9
2015年9月 JOURNAL OF UNIVERSITY OF SCIENCE AND TECHNOLOGY OF CHINA Sep. 2015

文章编号:0253-2778(2015)09-0709-08

一种生存数据非线性充分降维子空间估计的新方法

崔文泉,吴成龙

(中国科学技术大学管理学院统计与金融系,安徽合肥 230026)

摘要:给出了一种估计生存数据非线性充分降维子空间的新方法. 利用再生核 Hilbert 空间性质以及双切片思想,建立广义特征谱分解问题与获得充分降维子空间的联系,以此估计生存时间和删失时间的联合非线性降维中心子空间. 进一步结合 SDR 中心子空间的性质,通过联合 SDR 中心子空间来估计权重函数,在算法实现过程中,利用迭代思想,达到提高估计效率的目的. 最后通过数值模拟来说明该方法的优良性.

关键词:再生核 Hilbert 空间;充分降维;切片逆回归;生存数据

中图分类号:O212.7 **文献标识码**:A doi:10.3969/j.issn.0253-2778.2015.09.001

2010 Mathematics Subject Classification: Primary 62G08; Secondary 62N99

引用格式: Cui Wenquan, Wu Chenglong. An approach to estimating nonlinear sufficient dimension reduction subspace for censored survival data[J]. Journal of University of Science and Technology of China, 2015,45(9):709-716.
崔文泉,吴成龙. 一种生存数据非线性充分降维子空间估计的新方法[J]. 中国科学技术大学学报, 2015,45(9):709-716.

An approach to estimating nonlinear sufficient dimension reduction subspace for censored survival data

CUI Wenquan, WU Chenglong

(*Department of Statistics and Finance, School of Management, University of Science and of Technology of China, Hefei* 230026, *China*)

Abstract: An approach was proposed to estimating the nonlinear sufficient dimension reduction (SDR) subspace for survival data with censorship. Based on the theory of reproducing kernel Hilbert spaces (RKHS) and the double slicing procedure, the joint nonlinear sufficient dimension reduction central subspace was estimated by means of the generalized eigen-decomposition equation. And the weight function was estimated by the definition and property of SDR central subspace. The efficiency was improved by the iteration method while the algorithm was being implemented. Finally, the performance of the proposed method was illustrated on simulated data.

Key words: RKHS; sufficient dimension reduction; sliced inverse regression; survival data

---

收稿日期:2015-03-03;**修回日期**:2015-05-21

基金项目:安徽省自然科学基金(1308085MA02),中国科学院知识创新工程(KJCX3-SYW-S02)资助.

作者简介:崔文泉(通讯作者),男,1964年生,博士. 研究方向:数理统计. E-mail:wqcui@ustc.edu.cn

0 引言

充分降维(sufficient dimension reduction,SDR)是处理具有稀疏结构高维数据的一类重要方法. 对于响应变量 Y 和 p 维协变量 X 的关系,变量选择假设仅少数协变量对响应变量有影响,而充分降维假设 Y 依赖于协变量的 $q(q<p)$ 个线性组合,也即给定 $\boldsymbol{P}_{\mathscr{B}}X$,$Y$ 和 X 相互独立,表示为

$$X \perp Y \mid \boldsymbol{P}_{\mathscr{B}}X \tag{1}$$

式中,$\{\boldsymbol{\beta}_j\}_{j=1}^{q}$ 为充分降维方向,$\mathscr{B}=\mathrm{span}\{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_q\}$,是 SDR 中心子空间[1],为所有 SDR 子空间的交,$\boldsymbol{P}_{\mathscr{B}}$ 是空间 $\mathscr{B}$ 上正交投影,意义为 $\boldsymbol{P}_A=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}$ 是由 A 的列向量张成的线性子空间上的正交投影矩阵. 充分降维通过寻求 $\mathscr{B}$ 达到将 p 维输入降到具有 q 维特征结构的目的. 对于 $\mathscr{B}$ 的估计方法已有若干文献,读者可以参看综述文献[2].

具有删失机制的生存数据是统计学研究的一类重要数据类型. 该类数据的观测不完全性增大了对其进行充分降维研究的难度,使得相应理论方法的研究进展较为缓慢. Li 等[3]将切片逆回归[4](sliced inverse regression,SIR)推广到处理生存数据降维,该领域的其他若干研究可参阅文献[5-9]. 然而大多数已有的针对生存数据充分降维的研究都是在(1)框架下进行的,其结果只是获得协变量 X 的线性组合特征. 本文首次利用再生核 Hilbert 空间(reproducing kernel Hilbertspace,RKHS)理论基于核技术(kernel trick)提出处理生存数据非线性降维方法(RKHS-based double SIR,RDSIR),克服了当协变量对生存时间的影响具有非线性特征时非线性 SDR 中心子空间估计问题. RDSIR 假定生存时间 T 和 X 的关系满足如下模型:

$$T = f(u_1(X),\cdots,u_q(X),\epsilon) \tag{2}$$

式中,f 为未知函数,随机误差 ϵ 和 X 相互独立,$u_j(\cdot)\in\mathscr{H}_k$,$\mathscr{H}_k$ 是以 k 为再生核的 RKHS. 模型(2)是具有相当的普遍适用性,如,对于该模型在如下的特殊情形下,当模型(2)中的 $u_j(X)$ 具有更为简单线性结构时,即 $u_j(X)$ 是 X 的线性组合,则不难看出,常见的回归模型,如线性回归模型,广义线性模型,单指标模型,投影寻踪模型,甚至具有线性结构的神经网络模型,均可看成是此情形下该模型的特例.

对于本文的方法,首先通过适当建立的等距同构映射将对连续型无穷维函数的研究转化为对离散型数列函数的研究,并通过双切片思想来刻画删失信息. 进而得到生存时间 T 和删失时间 C 的联合非线性 SDR 中心子空间 $\mathscr{B}_J$ 的估计,进一步结合 SDR 中心子空间的性质,由 $\mathscr{B}_J$ 来估计权重函数,并通过广义特征分解问题得到生存时间的 SDR 中心子空间,在算法实现中,通过该上述步骤迭代,提高了估计精度. 在选取合适的再生核的情形下,RDSIR 对线性和非线性降维结构降维均有良好表现.

1 非线性充分降维方法

本文的非线性充分降维方法是建立在再生核 Hilbert 空间(RKHS)理论基础上的,这里首先对 RKHS 的相关内容进行简单的回顾和介绍,更多相关经典理论及核技术可参阅文献[12-13].

设 $\mathscr{H}_k$ 为定义在 $\mathscr{X}\subset\mathscr{R}^p$ 上的 Hilbert 空间,如果二元函数 $k(s,t)$,$s,t\in\mathscr{X}$ 满足如下条件:

① $k_s(t)=k(s,t)$ 是关于 t 的函数,并且满足 $k_s(t)\in\mathscr{H}_k$,对任意 $s\in\mathscr{X}$ 成立.

② 对任意 $s\in\mathscr{X}$,$f\in\mathscr{H}_k$,有

$$f(s) = \langle f,k_s\rangle \tag{3}$$

那么称 $\mathscr{H}_k$ 为再生核 Hilbert 空间,对应的 k 称为与 $\mathscr{H}_k$ 对应的再生核,其中(3)称为再生性质.

如果定义在 $\mathscr{X}$ 上的 Hilbert 空间 $\mathscr{H}$ 有再生核,那么该再生核 k 被 $\mathscr{H}$ 唯一确定. 若 $\mathscr{H}_k$ 是可分的 Hilbert 实函数空间,且有再生核 $k(s,t)$,$\{\phi_j\}_{j=1}^{\infty}$ 是 $\mathscr{H}_k$ 的标准正交基,那么 k 具有谱分解

$$k(s,t) = \sum_{j=1}^{\infty} a_j\phi_j(s)\phi_j(t),\ s,t\in\mathscr{X},$$

而且

$$\lim_{n\to\infty}\| k(s,t) - \sum_{j=1}^{n} a_j\phi_j(s)\phi_j(t)\| = 0.$$

定义映射 ϕ: $\mathscr{X}\mapsto l_2$,

$$\left.\begin{aligned}&\phi(s) = (\sqrt{a_1}\phi_1(s),\sqrt{a_2}\phi_2(s),\cdots)^{\mathrm{T}},\\&\langle\phi(s),\phi(t)\rangle = k(s,t),s,t\in\mathscr{X}\end{aligned}\right\} \tag{4}$$

在内积(4)下,空间 $\mathrm{span}\{\phi(s_1),\phi(s_2),\cdots\}$ 的完备化所对应的 Hilbert 空间记为 $\mathscr{H}$.

若有映射 $\mathscr{L}$: $\mathscr{L}k(s,\cdot)=\phi(s)$,满足

$$\langle\mathscr{L}k(s,\cdot),\mathscr{L}k(t,\cdot)\rangle_{\mathscr{H}_k} = k(s,t),$$

我们不难得,$\mathscr{L}$ 是 $\mathscr{H}_k$ 到 $\mathscr{H}$ 的等距同构映射.

易知 $\phi(X)$ 是取值于 $\mathscr{H}$ 的随机元. 假设 $\phi(X)$ 模的平方是关于 X 的概率分布几乎处处有界的,即

$$\|\phi(X)\|^2 = R(X,X) < \infty,\quad \text{a.s.} \tag{5}$$

一般地，对于取值于 $\mathscr{H}$ 的随机元 Z，假设 $E\|Z\|^2<\infty$，则 Z 的数学期望定义为 $E(Z)\in\mathscr{H}$，且满足 $\langle a,E(Z)\rangle=E\langle a,Z\rangle$，对所有 $a\in\mathscr{H}$. Z 的协方差算子定义为

$$\mathrm{Cov}(Z)=E[(Z-E(Z))\otimes(Z-E(Z))],$$

其中运算$\otimes$意义为：对于任意 $a,b\in\mathscr{H}$，有

$$(a\otimes b)f=\langle b,f\rangle a,$$

对任意 $f\in\mathscr{H}$.

设 T 为生存时间，C 为删失时间，X 为 p 维协变量. 记 $\Delta=I(T\leqslant C)$，$\widetilde{T}=T\wedge C$，假定在给定协变量 X 的条件下，T 与 C 相互独立，即 $T\perp C\mid X$. 样本观测数据 $\{\widetilde{T}_i,\Delta_i,X_i\}_{i=1}^n$ 来自 $(\widetilde{T},\Delta,X)$ 的样本容量为 n 的样本.

生存时间 T 与协变量 X 满足模型(2)，由 RKHS 再生性质及等距同构映射 $\mathscr{L}$，存在与 $u_j\in\mathscr{H}_k$ 是一一对应的 $\boldsymbol{\beta}_j\in\mathscr{H}$，使得

$$u_j(X)=\langle\boldsymbol{\beta}_j,\phi(X)\rangle \tag{6}$$

因此模型(2)可以表示为

$$T=g(\langle\boldsymbol{\beta}_1,\phi(X)\rangle,\cdots,\langle\boldsymbol{\beta}_q,\phi(X)\rangle,\epsilon) \tag{7}$$

称 $\{\boldsymbol{\beta}_j\}_{j=1}^q$ 是非线性 SDR 方向. 记 $\mathscr{B}_T$ 为生存时间 T 关于 X 的非线性 SDR 中心子空间，$\{\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_q\}$ 是 $\mathscr{B}_T$ 的标准正交基. 由 SDR 中心子空间的定义可知，给定 $\boldsymbol{P}_{\mathscr{B}_T}\phi(X)$，也即 $(u_1(X),\cdots,u_q(X))$，T 与 X 条件独立，表示为

$$T\perp X\mid\boldsymbol{P}_{\mathscr{B}_T}\phi(X) \tag{8}$$

式中，$\boldsymbol{P}_{\mathscr{B}_T}$ 是空间 $\mathscr{B}_T$ 上的投影算子.

由文献[11]，$\mathscr{B}_T$ 的基可通过如下的广义特征分解问题的非零特征值所对应的特征函数获得：

$$\Gamma\boldsymbol{\beta}=\lambda\Sigma\boldsymbol{\beta} \tag{9}$$

式中，Σ 是 $\phi(X)$ 的协方差算子，Γ 是 $\phi(X)$ 关于 T 的条件协方差算子，也即，

$$\Sigma=\mathrm{Cov}(\phi(X)),\ \Gamma=\mathrm{Cov}(E(\phi(X)\mid T)).$$

由(5)可得，Σ 是紧算子，其谱分解为

$$\Sigma=\sum_{j=1}^{\infty}\nu_j\psi_j\otimes\psi_j$$

$\{\nu_j\}_{j\geqslant1}$ 是非降的特征值序列，$\{\psi_j\}_{j\geqslant1}$ 是相应的特征函数. 在非线性降维特征的线性设计条件下 Γ 是紧算子，算子 Γ 的秩 $q_\Gamma\leqslant q$，其谱分解为

$$\Gamma=\sum_{j=1}^{q_\Gamma}\mu_j\varphi_j\otimes\varphi_j,$$

$\{\mu_j\}_{j\geqslant1}$ 是非降的特征值序列，$\{\varphi_j\}_{j\geqslant1}$ 是相应的特征函数. 所谓的非线性降维特征的线性设计条件，即对于任意 $\boldsymbol{\beta}\in\mathscr{H}$，存在常数 $c_0,c_1,\cdots,c_q$ 使得条件期望 $E(\langle\boldsymbol{\beta},\phi(X)\rangle\mid\langle\boldsymbol{\beta}_1,\phi(X)\rangle,\cdots,\langle\boldsymbol{\beta}_q,\phi(X)\rangle)$ 满足

$$E(\langle\boldsymbol{\beta},\phi(X)\rangle\mid\langle\boldsymbol{\beta}_1,\phi(X)\rangle,\cdots,\langle\boldsymbol{\beta}_q,\phi(X)\rangle)=c_0+c_1\langle\boldsymbol{\beta}_1,\phi(X)\rangle+\cdots+c_q\langle\boldsymbol{\beta}_q,\phi(X)\rangle.$$

设 $0=t_1<\cdots<t_L<t_{L+1}=\infty$ 为生存时间 T 取值范围的分割点，假设 $L>q$，分割区间 $\Pi_l=[t_l,t_{l+1})$. 假设

$$p_l=P(T\in\Pi_l)>0,\ l=1,\cdots,L.$$

记 $\boldsymbol{m}=E\{\phi(X)\}$，条件期望 $\boldsymbol{m}_h=E\{\phi(X)\mid T\in\Pi_l\}$，也即

$$\boldsymbol{m}_l=\frac{E\{\phi(X)1(T\in\Pi_l)\}}{P\{T\in\Pi_l\}}.$$

于是式(9)基于样本的经验广义特征分解问题为

$$\hat{\Gamma}\boldsymbol{\beta}=\lambda\hat{\Sigma}\boldsymbol{\beta} \tag{10}$$

式中，

$$\hat{\Gamma}=\sum_{l=1}^{L}\hat{p}_l(\hat{\boldsymbol{m}}_l-\hat{\boldsymbol{m}})\otimes(\hat{\boldsymbol{m}}_l-\hat{\boldsymbol{m}}) \tag{11}$$

式中，$\hat{p}_l,\hat{\boldsymbol{m}}_l,\hat{\boldsymbol{m}}$ 分别是 $p_l,\boldsymbol{m}_l,\boldsymbol{m}$ 的估计.

由引理 2.1，不难得到如下估计：

$$\hat{E}\{\phi(X)1(T\geqslant t)\}=\frac{1}{n}\sum_{i:\widetilde{T}_i\geqslant t}^{n}\phi(X_i)+\frac{1}{n}\sum_{i:\widetilde{T}_i<t,\Delta_i=0}^{n}\phi(X_i)\hat{w}(\widetilde{T}_i,t,X_i) \tag{12}$$

$$\hat{P}\{T\geqslant t\}=\frac{1}{n}\sum_{i=1}^{n}1(\widetilde{T}_i\geqslant t)+\frac{1}{n}\sum_{i:\widetilde{T}_i<t,\Delta_i=0}^{n}\hat{w}(\widetilde{T}_i,t,X_i) \tag{13}$$

记 $\hat{w}_{il}=\hat{P}(T_i\in\Pi_l\mid\Delta_i,\phi(X_i),\widetilde{T}_i)$，可得，

$$\hat{w}_{il}=\begin{cases}\hat{w}(\widetilde{T}_i,t_l,X_i)-\hat{w}(\widetilde{T}_i,t_{l+1},X_i), & \Delta_i=0;\\ I(t_l\leqslant\widetilde{T}_i<t_{l+1}), & \Delta_i=1.\end{cases}$$

于是由式(12)，(13)得

$$\hat{E}\{\phi(X)1(T\in\Pi_l)\}=\hat{E}\{\phi(X)1(T\geqslant t_l)\}-\hat{E}\{\phi(X)1(T\geqslant t_{l+1})\}=\frac{1}{n}\sum_{i=1}^{n}\phi(X_i)\hat{w}_{il} \tag{14}$$

$$\hat{P}(T\in\Pi_l)=\hat{P}(T\geqslant t_l)-\hat{P}(T\geqslant t_{l+1})=\frac{1}{n}\sum_{i=1}^{n}\hat{w}_{il} \tag{15}$$

记 $\Phi=(\phi(X_1),\cdots,\phi(X_n))$，将上述估计带入式(11)，进一步化简可得

$$\hat{\Gamma}=\frac{1}{n}\Phi J\Phi^{\mathrm{T}},$$

其中，$J=\frac{1}{n}\hat{W}\hat{P}^{-1}\hat{W}^{\mathrm{T}}-\boldsymbol{P}_{\mathbf{1}_n}$，$\boldsymbol{P}_A=A(A^{\mathrm{T}}A)^{-}A^{\mathrm{T}}$ 是向由 A 列向量张成的线性子空间上的正交投影矩阵，当 $A=\mathbf{1}_n$时，有 $\boldsymbol{P}_{\mathbf{1}_n}=\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}}$.

$$\hat{P}=\begin{pmatrix}\hat{p}_1 & & \\ & \ddots & \\ & & \hat{p}_L\end{pmatrix},\ \hat{W}=\begin{pmatrix}\hat{w}_{11} & \cdots & \hat{w}_{1L}\\ \vdots & \ddots & \vdots\\ \hat{w}_{n1} & \cdots & \hat{w}_{nL}\end{pmatrix}$$

对于式(10)特征函数求解，由定理1，该问题可以等价化为

$$\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha} \tag{16}$$

式中，$\boldsymbol{K}=(k(X_i,X_j))_{1\leqslant i,j\leqslant n}$为再生核$k$ 对应于样本的 Gram 矩阵，$D=\boldsymbol{I}_n-\boldsymbol{P}_{\mathbf{1}_n}$，为由 $\mathbf{1}_n$ 所张成线性子空间之正交补空间的投影矩阵. 为了解决过拟合(over-fitting)问题，使得算法实现具有稳定性，类似文献[14-16]，对式(16)进行正则化处理，在求解方程中加入正则化参数 τ，得到

$$\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda(\boldsymbol{K}D\boldsymbol{K}+n^2\tau\boldsymbol{I}_n))\boldsymbol{\alpha} \tag{17}$$

式中，τ 为 tuning 参数，可由交叉验证等方法进行选优.

下面讨论如何利用删失信息提高权函数 w 的估计效率. 为刻画删失信息的降维空间，假设删失时间 C 与 X 的函数关系可以表示为

$$C=g(v_1(X),\cdots,v_c(X),\varepsilon) \tag{18}$$

式中，g 为未知函数，$v_j(\cdot)\in\mathscr{H}_k$，随机误差 ε 和 X 独立. 类似式(7)，模型(18)可以表示为

$$C=g(\langle\boldsymbol{\gamma}_1,\phi(X)\rangle,\cdots,\langle\boldsymbol{\gamma}_c,\phi(X)\rangle,\varepsilon) \tag{19}$$

式中，$\boldsymbol{\gamma}_j\in\mathscr{H},j=1,\cdots,c$. 令 $\mathscr{B}_C$ 为删失时间C 关于X 的非线性 SDR 中心子空间，由 $\mathscr{B}_T$ 和 $\mathscr{B}_C$ 的并所得到的线性子空间称为非线性联合 SDR 中心子空间，记为 $\mathscr{B}_J$，其维数不超过 $q+c$. 给定 $\boldsymbol{P}_{\mathscr{B}_J}\phi(X)$，$(T,C)$和 X 相互独立，不难得到

$$w(t',t,X)=w(t',t,\boldsymbol{P}_{\mathscr{B}_J}\phi(X)).$$

由文献[11]，在线性设计条件下，由

$$E(\phi(X)\mid\widetilde{T},\Delta)=E(E(\phi(X)\mid T,C)\mid\widetilde{T},\Delta),$$

可得

$$E(\phi(X)\mid\widetilde{T},\Delta)-E(\phi(X))\in \operatorname{span}\{\Sigma\boldsymbol{\beta}_1,\cdots,\Sigma\boldsymbol{\beta}_q,\Sigma\boldsymbol{\gamma}_1,\cdots,\Sigma\boldsymbol{\gamma}_c\}.$$

在理论上，$\mathscr{B}_J=\operatorname{span}\{\boldsymbol{\beta}_{J1},\cdots,\boldsymbol{\beta}_{Jd}\}$可由算子情形下的如下广义特征分解的方程获得

$$\Gamma_{\eta_d}\boldsymbol{\beta}=\lambda\Sigma\boldsymbol{\beta} \tag{20}$$

式中，

$$\Gamma_{\eta_d}=\operatorname{Cov}(\eta_d(\widetilde{T},\Delta)),$$
$$\eta_d(\widetilde{T},\Delta)=E(\phi(X)\mid\widetilde{T},\Delta).$$

对于Γ_{η_d}的经验估计借助由双切片思想[3]获得，记$(\widetilde{T},\Delta)$取值的分割点为 $0=t_{\delta1}<\cdots<t_{\delta L_\delta}<t_{\delta(L_\delta+1)}=\infty,\delta=0,1$，对于删失和未删失样本，分割份数 L_0 和 L_1. 分割区间 $\Pi_{\delta l}=[t_{\delta l},t_{\delta(l+1)})$，$\hat{p}_{\delta l}=\frac{n_{\delta l}}{n}$，$n_{\delta l}=\#\{\Delta_i=\delta,\widetilde{T}_i\in\Pi_{\delta l}:i=1,\cdots,n\}$. 得到

$$\hat{\Gamma}_{\eta_d}=\sum_{\delta=0,1}\sum_{l=1}^{L_\delta}\hat{p}_{\delta l}(\hat{\boldsymbol{m}}_{\delta l}-\hat{\boldsymbol{m}})\otimes(\hat{\boldsymbol{m}}_{\delta l}-\hat{\boldsymbol{m}})^{\mathrm{T}}.$$

其中，

$$\hat{\boldsymbol{m}}_{\delta l}=(n\hat{p}_{\delta l})^{-1}\sum_{i=1}^{n}\phi(X_i)I(\Delta_i=\delta,\widetilde{T}_i\in\Pi_{\delta l}).$$

进一步整理化简可得

$$\hat{\Gamma}_{\eta_d}=\frac{1}{n}\Phi J_d\Phi^{\mathrm{T}} \tag{21}$$

式中，

$$J_d=BAB^{\mathrm{T}}-\boldsymbol{P}_{\mathbf{1}_n},$$
$$A=\operatorname{diag}(\boldsymbol{P}_{\mathbf{1}_{n_{01}}},\cdots,\boldsymbol{P}_{\mathbf{1}_{n_{0L_0}}},\boldsymbol{P}_{\mathbf{1}_{n_{11}}},\cdots,\boldsymbol{P}_{\mathbf{1}_{n_{1L_1}}}),$$

B 是 n 阶初等矩阵，使得 $\widetilde{\Phi}=\Phi B$，

$$\widetilde{\Phi}=(\widetilde{\Phi}^{(01)},\cdots,\widetilde{\Phi}^{(0L_0)},\widetilde{\Phi}^{(11)},\cdots,\widetilde{\Phi}^{(1L_1)}).$$

进而可得式(20)的经验版本：

$$\hat{\Gamma}_{\eta_d}\boldsymbol{\beta}=\lambda\hat{\Sigma}\boldsymbol{\beta} \tag{22}$$

式(22)的所有非零特征值所对应的特征函数$\{\hat{\boldsymbol{\beta}}_{J1},\cdots,\hat{\boldsymbol{\beta}}_{Jd}\}$为一组基，可得联合 SDR 中心子空间的估计. 由式(21)及 $\hat{\boldsymbol{\beta}}_{Jj}=\Phi\hat{\boldsymbol{\alpha}}_{Jj}(j=1,\cdots,d)$，具有无穷维算子运算结构的式(22)有类似于式(16)的有限维形式 $\boldsymbol{K}J_d\boldsymbol{K}\boldsymbol{\alpha}=\lambda\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha}$，在求解过程中，使用相应的正则化方法，即：$\boldsymbol{K}J_d\boldsymbol{K}\boldsymbol{\alpha}=\lambda(\boldsymbol{K}D\boldsymbol{K}+n^2\tau\boldsymbol{I}_n)\boldsymbol{\alpha}$.
由此获得 d 个降维特征的估计为

$$\boldsymbol{K}(X_{\leqslant n},X)^{\mathrm{T}}\hat{\boldsymbol{\alpha}}_{Jj},j=1,\cdots,d,$$

其中，$\boldsymbol{K}(X_{\leqslant n},X)=(k(X_1,X),\cdots,k(X_n,X))^{\mathrm{T}}$.

由引理2.1中 w 的表达，可通过求得 Λ 的估计进而获得 w 的估计. 利用所求的 $\mathscr{B}_J$ 的估计得到 $\hat{\Lambda}$ 如下：

$$\hat{\Lambda}(t',t\mid\boldsymbol{P}_{\mathscr{B}_J}\phi(X))=$$

$$\frac{1}{n}\sum_{i:t'<\widetilde{T}_i<t,\Delta_i=1}^{n}\frac{H_d(h_n^{-1}(\boldsymbol{K}(X_{\leqslant n},X_i)-\boldsymbol{K}(X_{\leqslant n},X))^{\mathrm{T}}\hat{\boldsymbol{\alpha}}_J)}{h_n^d\hat{S}_{\widetilde{T}}\widetilde{T}_i\mid\boldsymbol{P}_{\mathscr{B}_J}\phi(X_i)\hat{f}(\boldsymbol{P}_{\mathscr{B}_J}\phi(X))},$$

$$\hat{S}_{\widetilde{T}}(\widetilde{T}_i\mid\boldsymbol{P}_{\mathscr{B}_J}\phi(X_i))=$$

$$\frac{1}{n}\sum_{j:\tilde{T}_j>\tilde{T}_i}^{n}\frac{h_n^{-d}H_d(h_n^{-1}(\boldsymbol{K}(X_{\leqslant n},X_j)-\boldsymbol{K}(X_{\leqslant n},X_i))^{\mathrm{T}}\hat{\boldsymbol{\alpha}}_J)}{\hat{f}(\boldsymbol{P}_{\mathscr{B}_J}\phi(X_i))},$$

$$\hat{f}(\boldsymbol{P}_{\mathscr{B}_J}\phi(X))=$$

$$\frac{1}{n}\sum_{j=1}^{n}h_n^{-d}H_d(h_n^{-1}(\boldsymbol{K}(X_{\leqslant n},X_j)-\boldsymbol{K}(X_{\leqslant n},X))^{\mathrm{T}}\hat{\boldsymbol{\alpha}}_J),$$

式中，$H_d(\cdot)$是定义在 $\mathscr{R}^d$ 上的核函数，一般取乘积形式 $H_d(X)=H(X_{(1)})\cdots H(X_{(d)})$.

由上述 Λ 的估计可得 w 的估计：

$$\hat{w}(t',t,\boldsymbol{P}_{\mathscr{B}_J}\phi(X))=\exp\{-\hat{\Lambda}(t',t\mid\boldsymbol{P}_{\mathscr{B}_J}\phi(X))\},$$

进而由式(14)，(15) 及 $\hat{W}$ 和 $\hat{P}$ 的表达求得 J.

由式(17)求得单位特征向量记为 $\hat{\boldsymbol{\alpha}}_j^0$，由 $\boldsymbol{\beta}_j$ 是 $\mathscr{H}$ 中的单位特征函数，那么$\langle\boldsymbol{\beta}_j,\boldsymbol{\beta}_j\rangle=1$，于是

$$\langle\Phi\boldsymbol{\alpha}_j,\Phi\boldsymbol{\alpha}_j\rangle=\boldsymbol{\alpha}_j^{\mathrm{T}}\mathbf{K}\boldsymbol{\alpha}_j=1.$$

令

$$\boldsymbol{\alpha}_j=\boldsymbol{\alpha}_j^0\sqrt{\boldsymbol{\alpha}_j^{0\mathrm{T}}\mathbf{K}\boldsymbol{\alpha}_j^0}$$

便可实现 $\boldsymbol{\beta}_j$ 单位化.

2　若干理论结果

为了求解 m_j 与权函数 w 的估计，首先给出如下引理 2.1.

引理 2.1　对于 $\phi(X)\in\mathscr{H}$，由式(8)，如下等式成立：

$$E[\phi(X)I(T\geqslant t)]=E[\phi(X)I(\tilde{T}\geqslant t)]+E[\phi(X)I(\tilde{T}<t,\Delta=0)w(\tilde{T},t,\phi(X)],$$

$$E[I(T\geqslant t)]=E[I(\tilde{T}\geqslant t)]+E[I(\tilde{T}<t,\Delta=0)w(\tilde{T},t,\phi(X))],$$

权重函数

$$w(t',t,\phi(X))=\exp\{-\Lambda(t',t\mid\phi(X))\},$$

且

$$\Lambda(t',t\mid\phi(X))=E\left\{\frac{I(t'\leqslant\tilde{T}<t,\Delta=1)}{S_{\tilde{T}}(\tilde{T}\mid\phi(X))}\middle|\phi(X)\right\}.$$

引理 2.1 的证明见文献[10]，这里略过. 下面我们给出引理 2.2，其建立了一种由有限维情形的特征分解问题获得无穷维广义特征分解问题的求解方式，并进一步得到特征 u_j 估计的表示.

引理 2.2　无穷维情形下经验广义特征分解式(10)中 $\boldsymbol{\beta}$ 的解具有如下表达形式：

$$\hat{\boldsymbol{\beta}}_j=\sum_{i=1}^{n}\hat{\boldsymbol{\alpha}}_{j(i)}\phi(X_i)=\Phi\hat{\boldsymbol{\alpha}}_j,\ j=1,\cdots,q\tag{23}$$

式中，$\hat{\boldsymbol{\alpha}}_1,\cdots,\hat{\boldsymbol{\alpha}}_q$ 由如下有限维情形下经验广义特征分解求得：

$$\boldsymbol{KJK\alpha}=\lambda\boldsymbol{KDK\alpha}\tag{24}$$

且式(10)与式(24)是等价的. 对于正则化分解

$$\hat{\Sigma}\hat{\Gamma}\boldsymbol{\beta}=\lambda(\hat{\Sigma}^2+\tau I)\boldsymbol{\beta}\tag{25}$$

有类似的结论成立，其中 I 是恒等算子，即式(25)中 $\boldsymbol{\beta}$ 的解也有式(23)的形式，其中的 $\hat{\boldsymbol{\alpha}}_j$ 是$\boldsymbol{KJK\alpha}=\lambda(\boldsymbol{KPK}+n^2\tau\boldsymbol{I}_n)\boldsymbol{\alpha}$ 的解.

引理 2.2 的结果可以推广到联合 SDR 问题等情形，如当将上述的 J 换成 J_d时，也有相应的结论成立. 由式(6)及引理 2.2 不难得到如下的表示定理.

定理 2.3　设样本观测数据为$\{\tilde{T}_i,\Delta_i,X_i\}_{i=1}^n$，则模型(2)中特征 u_j 的估计可表示为

$$\hat{u}_j(X)=\boldsymbol{K}(X_{\leqslant n},X)\hat{\boldsymbol{\alpha}}_j,\ j=1,\cdots,q,$$

其中，$\hat{\boldsymbol{\beta}}_j=\Phi\hat{\boldsymbol{\alpha}}_j$.

下面给出本文方法的渐近性质. 设定义在可分的 Hilbert 空间 $\mathscr{H}$ 上的线性算子 $\mathscr{L}$ 满足下式，

$$\|\mathscr{L}\|_{HS}^2=\sum_{i=1}^{\infty}\|\mathscr{L}\psi_i\|_{\mathscr{H}}^2<\infty,$$

其中，$\{\psi_i\}$是标准正交基，那么 $\mathscr{L}$ 属于 Hilbert-Schmidt 类. Hilbert-Schmidt 类张成一个新的 Hilbert 空间，并以$\|\cdot\|_{HS}$为范数. 如果 $\mathscr{S}$为 $\mathscr{H}$上的有界算子，那么 $\mathscr{SL}$ 和 $\mathscr{LS}$ 都属于 Hilbert-Schmidt 类，并且有

$$\|\mathscr{SL}\|_{HS}\leqslant\|\mathscr{S}\|\ \|\mathscr{L}\|_{HS},$$
$$\|\mathscr{LS}\|_{HS}\leqslant\|\mathscr{L}\|\ \|\mathscr{S}\|_{HS},$$

其中，$\|\cdot\|$是算子范数

$$\|\mathscr{L}\|^2=\sup_{f\in\mathscr{H}}\frac{\|\mathscr{L}f\|^2}{\|f\|^2}.$$

为研究式(17)所得估计量的相合性等性质，本文给出如下假设：存在正数 $\kappa\in(0,1/2]$，使得

(C1)　$|\hat{p}_l-p_l|=O_p(n^{-\kappa})$；

(C2)　$\left\|\frac{1}{n}\Phi\hat{\omega}_l-E[\phi(X)I(T\in\Pi_l)]\right\|_{HS}=O_p(n^{-\kappa})$.

定理 2.4　设 $ER(X,X)^2<\infty$，假设(C1)～(C2)成立，那么对于任意给定的正整数 N，有下式成立：

$$\|(\hat{\Sigma}^2+\tau I)^{-1}\hat{\Sigma}\hat{\Gamma}-\Sigma^{-1}\Gamma\|_{HS}=O_p\left(\frac{1}{\tau n^{\kappa}}\right)+\sum_{j=1}^{q_\Gamma}\left(\frac{\tau}{\nu_N^2}\|\Psi_N(\tilde{\varphi}_j)\|+\|\Psi_N^{\perp}(\tilde{\varphi}_j)\|\right)\tag{26}$$

式中，$\tilde{\varphi}_j=\Sigma^{-1}\varphi_j$；$\Psi_N$ 和 $\Psi_N^{\perp}$ 是互为正交空间的投影算子，

$$\Psi_N=\sum_{j=1}^{N}\psi_j\otimes\psi_j,\ \Psi_N^{\perp}=\sum_{j=N+1}^{\infty}\psi_j\otimes\psi_j.$$

如果光滑参数 τ 满足 $\lim\limits_{n\to\infty}\tau=0$，$\lim\limits_{n\to\infty}\tau n^{\kappa}=\infty$，则

$$\|(\hat{\Sigma}^2+\tau I)^{-1}\hat{\Sigma}\hat{\Gamma}-\Sigma^{-1}\Gamma\|_{HS}=o_p(1) \tag{27}$$

定理 2.5 设 $ER(X,X)^2<\infty$，假设(C1)~(C2)成立，且 $\lim\limits_{n\to\infty}\tau=0$，$\lim\limits_{n\to\infty}\tau n^{\kappa}=\infty$，那么

$$|\langle\hat{\boldsymbol{\beta}}_j,\phi(X)\rangle-\langle\boldsymbol{\beta}_j,\phi(X)\rangle|=o_p(1),\ j=1,\cdots,q_{\Gamma} \tag{28}$$

式中，$\{\boldsymbol{\beta}_j\}_{j=1}^{q_{\Gamma}}$ 为式(9)的非零特征值所对应的特征函数，$\{\hat{\boldsymbol{\beta}}_j\}_{j=1}^{q_{\Gamma}}$ 是相应的式(25)的特征函数.

进一步的，如果 $\{\boldsymbol{\beta}_j\}_{j=1}^{q_{\Gamma}}$ 仅仅与 Σ 的有限个特征函数有关，那么收敛速度为 $O(n^{-\kappa/2})$，即

$$\left.\begin{array}{c}|\langle\hat{\boldsymbol{\beta}}_j,\phi(X)\rangle-\langle\boldsymbol{\beta}_j,\phi(X)\rangle|=O_p(n^{-\kappa/2}),\\ j=1,\cdots,q_{\Gamma}\end{array}\right\} \tag{29}$$

定理 2.4 和定理 2.5 的证明本文从略. 我们不难推得，定理 2.5 可以看成是定理 2.4 的推论，当式(27)成立时，由算子扰动理论可得式(28)成立. 如果 $\{\boldsymbol{\beta}_j\}_{j=1}^{q_{\Gamma}}$ 仅仅与 Σ 的有限个特征函数有关时，则存在正整数 N 使得

$$\mathscr{D}=\operatorname{span}\{\Sigma\psi_i,i=1,\cdots,N\},$$

因此有 $\tilde{\varphi}_j=\Sigma^{-1}(\mathscr{D})\subset\operatorname{span}\{\psi_i,i=1,\cdots,N\}$，$\Psi_N^{\perp}(\tilde{\varphi}_j)=0$. 当取 $\tau=n^{-\kappa/2}$ 时，可得式(29).

引理 2.2 证明 设 β 为满足特征方程式(10)的特征函数，由 $\hat{\Sigma}=\frac{1}{n}\Phi D\Phi^{\mathrm{T}}$，$\hat{\Gamma}=\frac{1}{n}\Phi J\Phi^{\mathrm{T}}$，将 β 分解成 $\beta=\beta_1+\beta_2$，其中，$\beta_1\in\mathscr{H}_{\Phi}=\operatorname{span}\{\Phi\}$，$\beta_2\in\mathscr{H}_{\Phi}^{\perp}$，由此

$$\hat{\Sigma}\beta_2=0,\ \hat{\Gamma}\beta_2=0,$$

而 SDR 中心子空间是所有 SDR 子空间的交，因此式(10)的对应于非零特征值的特征函数有式(23)的形式.

设再生核 k 的非零特征值的个数为 $d_k\leqslant\infty$，则 $\phi(s)=(\sqrt{a_1}\phi_1(s),\sqrt{a_2}\phi_2(s),\cdots,\phi_{d_k}(s))^{\mathrm{T}}$.

(Ⅰ) 当 $d_k<\infty$ 时，即 $\mathscr{H}$ 为有限维，此时 Φ 是 $d_k\times n$ 的矩阵，设 Φ 有如下奇异值分解：

$$\Phi=UDV=$$

$$(u_1,\cdots,u_{d_k})\begin{pmatrix}\overline{D}_{d\times d} & 0_{d\times(n-d)}\\ 0_{(d_k-d)}\times d & 0_{(n-d_k)\times(n-d)}\end{pmatrix}\begin{bmatrix}v_1^{\mathrm{T}}\\ \vdots\\ v_n^{\mathrm{T}}\end{bmatrix}=$$

$$\overline{U}\,\overline{D}\,\overline{V}^{\mathrm{T}} \tag{30}$$

式中，$\overline{U}=(u_1,\cdots,u_d)$，$\overline{V}=(v_1,\cdots,v_d)$，$\overline{D}=\overline{D}_{d\times d}$ 为对角矩阵，且 $\overline{U}^{\mathrm{T}}\overline{U}=\overline{V}^{\mathrm{T}}\overline{V}=I_d$.

($\Rightarrow$) 由 $\hat{\Gamma}\boldsymbol{\beta}=\lambda\hat{\Sigma}\boldsymbol{\beta}$ 成立，那么对于 $i=1,\cdots,n$ 有 $\langle\phi(X_i),\hat{\Gamma}\boldsymbol{\beta}\rangle=\langle\phi(X_i),\lambda\hat{\Sigma}\boldsymbol{\beta}\rangle$，可得

$$\Phi^{\mathrm{T}}\left(\frac{1}{n}\Phi J\Phi^{\mathrm{T}}\boldsymbol{\beta}\right)=\frac{1}{n}\Phi^{\mathrm{T}}\Phi J\Phi^{\mathrm{T}}\Phi\boldsymbol{\alpha}=\frac{1}{n}\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha},$$

$$\Phi^{\mathrm{T}}\left(\frac{1}{n}\lambda\Phi D\Phi^{\mathrm{T}}\boldsymbol{\beta}\right)=\frac{1}{n}\lambda\Phi^{\mathrm{T}}\Phi D\Phi^{\mathrm{T}}\Phi\boldsymbol{\alpha}=\lambda\frac{1}{n}\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha},$$

因此有 $\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha}$.

($\Leftarrow$) 由 $\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha}$ 成立，且有 $\boldsymbol{\beta}=\Phi\boldsymbol{\alpha}$，$\boldsymbol{K}=\Phi^{\mathrm{T}}\Phi=\overline{V}\,\overline{D}^2\overline{V}^{\mathrm{T}}$，故

$$\begin{aligned}&\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda\boldsymbol{K}D\boldsymbol{K}\boldsymbol{\alpha}\Rightarrow\\ &\quad\overline{V}\,\overline{D}^2\overline{V}^{\mathrm{T}}J\Phi^{\mathrm{T}}\Phi\boldsymbol{\alpha}=\lambda\overline{V}\,\overline{D}^2\overline{V}^{\mathrm{T}}D\Phi^{\mathrm{T}}\Phi\boldsymbol{\alpha}\\ &\overset{\Delta}{\Rightarrow}\overline{U}\,\overline{D}\,\overline{V}^{\mathrm{T}}J\Phi^{\mathrm{T}}\boldsymbol{\beta}=\lambda\overline{U}\,\overline{D}\,\overline{V}^{\mathrm{T}}D\Phi^{\mathrm{T}}\boldsymbol{\beta}\Rightarrow\\ &\Phi J\Phi^{\mathrm{T}}\boldsymbol{\beta}=\lambda\Phi D\Phi^{\mathrm{T}}\boldsymbol{\beta}\Rightarrow\\ &\hat{\Gamma}\boldsymbol{\beta}=\lambda\hat{\Sigma}\boldsymbol{\beta}\end{aligned}$$

其中第 $\overset{\Delta}{\Rightarrow}$ 步，在等式两边分别乘上 $\overline{U}\,\overline{D}^{-1}\overline{V}^{\mathrm{T}}$ 即可得到.

(Ⅱ) 当 $d_k=\infty$ 时，此时 $\mathscr{H}$ 为无穷维，可以用如下方式定义 Φ，Φ^{T} 以及 Φ 的奇异值分解. Φ 是一个从 $\mathscr{R}^n$ 到 $\mathscr{H}$ 的算子：$\Phi\boldsymbol{\alpha}=\sum\limits_{i=1}^{n}\boldsymbol{\alpha}_{(i)}\phi(X_i)$，$\boldsymbol{\alpha}\in\mathscr{R}^n$，且 Φ^{T} 为其共轭算子：

$$\Phi^{\mathrm{T}}\boldsymbol{\beta}=(\langle\phi(X_1),\boldsymbol{\beta}\rangle,\cdots,\langle\phi(X_n),\boldsymbol{\beta}\rangle)^{\mathrm{T}},\ \boldsymbol{\beta}\in\mathscr{H},$$

由于 $\hat{\Sigma}$ 的秩 $d\leqslant n$，有如下分解

$$\hat{\Sigma}=\sum_{i=1}^{d_k}\sigma_i v_i\otimes v_i=\sum_{i=1}^{d}\sigma_i v_i\otimes v_i \tag{31}$$

式中，$\sigma_1\geqslant\cdots\geqslant\sigma_d\geqslant\sigma_{d+1}=\cdots=0$，记 $\overline{U}=(v_1,\cdots,v_d)$，类似对 Φ 处理，可以定义算子 $\overline{U}$ 及其共轭算子 $\overline{U}^{\mathrm{T}}$. 由式(31)可得 $\phi(X_i)\in\operatorname{span}\{v_1,\cdots,v_d\}$，记 $\phi(X_i)=\overline{U}\omega_i$，$\Omega=(\omega_1,\cdots,\omega_n)^{\mathrm{T}}$，易得 $\Omega^{\mathrm{T}}\Omega=\operatorname{diag}(n\sigma_1,\cdots,n\sigma_d)$，令

$$\overline{D}_{d\times d}=\operatorname{diag}(\sqrt{n\sigma_1},\cdots,\sqrt{n\sigma_d}),\overline{V}=\Omega\overline{D}^{-1},$$

可得算子 Φ 的奇异值分解

$$\Phi=\overline{U}\,\overline{D}\,\overline{V}^{\mathrm{T}}.$$

后续推导类似情形(Ⅰ).

表 1　具有非线性降维结构时 RDSIR 和 DSIR 的降维表现：Acc 均值(标准差)

Tab. 1　The nonlinear dimension reduce results of RDSIR and DSIR: mean (standard deviation)

n	censored percent	Acc1		Acc2		Acc3	
		RDSIR①	DSIR②	RDSIR③	DSIR④	RDSIR⑤	DSIR⑥
100	0%	0.496(0.839)	0.888(0.611)	0.638(0.808)	0.064(0.581)	0.811(0.763)	0.577(0.642)
	20%	0.503(1.018)	0.844(1.419)	0.625(1.011)	0.070(0.424)	0.801(0.591)	0.555(1.185)
	40%	0.465(1.099)	0.841(0.735)	0.616(1.056)	0.065(0.469)	0.771(0.604)	0.555(0.870)
	60%	0.379(1.151)	0.700(1.577)	0.601(1.252)	0.071(0.426)	0.702(0.843)	0.459(1.171)
200	0%	0.612(0.773)	0.958(0.176)	0.694(0.621)	0.063(0.361)	0.937(0.173)	0.623(0.517)
	20%	0.597(0.724)	0.945(0.241)	0.707(0.438)	0.051(0.383)	0.924(0.221)	0.618(0.528)
	40%	0.557(0.853)	0.935(0.313)	0.705(0.475)	0.055(0.405)	0.896(0.295)	0.610(0.574)
	60%	0.487(0.960)	0.876(0.845)	0.683(0.671)	0.061(0.469)	0.834(0.464)	0.573(0.819)

由(Ⅰ)和(Ⅱ)即得式(10)与式(24)解的等价性，并且上述推导过程不难推广到正则化情形，从而完成引理的证明. □

3　实现算法及模拟分析

本小节首先给出 $\mathscr{B}_T$ 估计的具体算法实现，然后通过模拟研究显示方法在小样本情形下的良好表现.

Step 0　初始化：给定分割份数 $L(L>q)$，迭代次数 $\nu=1$ 以及再生核 k.

Step 1　由样本及 k 计算得到 Gram 矩阵. 利用双切片思想估计 J_d，求解

$$\boldsymbol{K}J_d\boldsymbol{K}\boldsymbol{\alpha}=\lambda(\boldsymbol{K}\boldsymbol{D}\boldsymbol{K}+n^2\tau\boldsymbol{I}_n)\boldsymbol{\alpha}$$

得到特征向量矩阵 $\boldsymbol{\alpha}_J^0=(\boldsymbol{\alpha}_{J1}^0,\cdots,\boldsymbol{\alpha}_{Jd}^0)$，单位化得到 $\boldsymbol{\alpha}_J=(\boldsymbol{\alpha}_{J1},\cdots,\boldsymbol{\alpha}_{Jd})$，其中,

$$\boldsymbol{\alpha}_{Jj}=\boldsymbol{\alpha}_{Jj}^0/\sqrt{\boldsymbol{\alpha}_{Jj}^{0\mathrm{T}}\boldsymbol{K}\boldsymbol{\alpha}_{Jj}^0},$$

得到 $\mathscr{B}_J=\Phi\boldsymbol{\alpha}_J$，由

$$w(t',t,X)=w(t',t,\boldsymbol{P}_{\mathscr{B}_J}\phi(X))$$

得到权重矩阵 $\hat{W}$ 估计. 进而由

$$\boldsymbol{K}J\boldsymbol{K}\boldsymbol{\alpha}=\lambda(\boldsymbol{K}\boldsymbol{D}\boldsymbol{K}+n^2\tau\boldsymbol{I}_n)\boldsymbol{\alpha}\tag{32}$$

特征分解并单位化得到 $\boldsymbol{\alpha}_T^{(\nu)}=(\boldsymbol{\alpha}_{T1}^{(\nu)},\cdots,\boldsymbol{\alpha}_{Tq}^{(\nu)})$，得到估计 $\mathscr{B}_T^{(\nu)}=\Phi\boldsymbol{\alpha}_T^{(\nu)}$.

Step 2　进一步利用 $\mathscr{B}_T^{(\nu)}$ 将 $\mathscr{B}_J$ 更新为 $\mathscr{B}_{J(T)}^{(\nu)}$，以及

$$w(t',t,X)=w(t',t,\boldsymbol{P}_{\mathscr{B}_{J(T)}}^{(\nu)}\phi(X))$$

得到 $\hat{W}$ 新的估计. 再次通过式(32)计算得到 $\mathscr{B}_T^{(\nu+1)}$.

Step 3　令 $\nu=\nu+1$，重复 Step 2 直至收敛. 本文以

$$\boldsymbol{\alpha}_T^{(\nu+1)}(\boldsymbol{\alpha}_T^{(\nu+1)\mathrm{T}}\boldsymbol{\alpha}_T^{(\nu+1)})^{-1}\boldsymbol{\alpha}_T^{(\nu+1)\mathrm{T}}-\boldsymbol{\alpha}_T^{(\nu)}(\boldsymbol{\alpha}_T^{(\nu)\mathrm{T}}\boldsymbol{\alpha}_T^{(\nu)})^{-1}\boldsymbol{\alpha}_T^{(\nu)\mathrm{T}}$$

的最大奇异值作为是否收敛的判别值，例如当其小于 1×10^{-4} 时，令此时 $\mathscr{B}_T^{(\nu)}$ 为 $\mathscr{B}_T$ 的最终估计.

下面通过数值模拟来研究 RDSIR 的表现，考虑生存时间满足比例失效模型 PH(proportional hazards model)[17]，失效函数为

$$\lambda(t\mid X)=\lambda_0(t)\exp\{u_1(X)\}$$

具有非线性降维结构，$\lambda_0(t)=\lambda_0$ 为常数.

模拟生存时间由式(33)生成(见文献[10])，删失时间 C 服从均匀分布 $U(0,c)$，通过 c 的大小调节删失比例.

$$T=-\frac{\ln U}{\exp\{u_1(X)\}}\tag{33}$$

U 服从均匀分布 $U(0,1)$，$p=10$ 维协变量 X 服从多元正态分布 $N(0,\ 0.8^2I_{10})$,

$$u_1(X)=u_1^{(1)}(X)+u_1^{(2)}(X),$$

$$u_1^{(1)}(X)=\sum_{i=1}^{5}X_{(i)},u_1^{(2)}(X)=\sum_{i=6}^{10}X_{(i)}^2,$$

可以看出 $u_1^{(1)}(X)$ 和 $u_1^{(2)}(X)$ 分别是线性和非线性降维结构，模拟中取 $\tilde{\tau}=100$，再生核

$$k(s,t)=(\langle s,t\rangle+1)^2,$$

训练样本(training)为 $n_{\mathrm{tr}}=100,200$，测试样本(test)为 $n_{\mathrm{te}}=200$. 记 $\widetilde{X}=(u_1(X),\cdots,u_q(X))$，定义 $u_1(\cdot)$ 基于样本的估计精度(accuracy)为

$$\mathrm{Acc}(\hat{u}_1(X))=\max_{\boldsymbol{b}\in\mathscr{R}^q}\rho(\hat{u}_1(X),\boldsymbol{b}^{\mathrm{T}}\widetilde{X})=$$

$$\rho(\hat{u}_1(X),u_1(X)),\ q=1.$$

RDSIR 为本文方法，DSIR 为文献[3]提出的估计方法，当删失比例为 0%，20%，40%，60% 时分别用这两种方法处理该数据并进行比较，结果见表 1. 分别考虑精度

$$\mathrm{Acc1}=\rho(\hat{u}_1(X),u_1^{(1)}(X)),$$

$$\mathrm{Acc2}=\rho(\hat{u}_1(X),u_1^{(2)}(X)),$$
$$\mathrm{Acc3}=\rho(\hat{u}_1(X),u_1(X)).$$

从表1可以看出,DSIR能够获取线性降维结构部分(②列),而对于非线性部分获取信息较少(④列),相比而言RDSIR能够获取较多的非线性部分的信息(⑤列),并且包含了线性降维结构信息.

参考文献(References)

[1] Cook R D. Regression Graphics: Ideas for Studying Regressions Through Graphics[M]. New York: Wiley, 1998: 215.

[2] Ma Y, Zhu L. A review on dimension reduction[J]. International Statistical Review, 2013, 81(1): 134-150.

[3] Li K C, Wang J L, Chen C H. Dimension reduction for censored regression data[J]. The Annals of Statistics, 1999, 27(1): 1-23.

[4] Li K C. Sliced inverse regression for dimension reduction[J]. Journal of the American Statistical Association, 1991, 86(414):316-327.

[5] Li L, Li H. Dimension reduction methods for microarrays with application to censored survival data[J]. Bioinformatics, 2004, 20(18): 3 406-3 412.

[6] Shevlyakova M, Morgenthaler S. Sliced inverse regression for survival data[J]. Statistical Papers, 2014, 55(1):209-220.

[7] Wen X M. On sufficient dimension reduction for proportional censorship model with covariates[J]. Computational Statistics & Data Analysis, 2010, 54(8): 1 975-1 982.

[8] Xia Y, Zhang D, Xu J. Dimension reduction and semiparametric estimation of survival models[J]. Journal of the American Statistical Association, 2010, 105(489): 278-290.

[9] Lu W, Li L. Sufficient dimension reduction for censored regressions[J]. Biometrics, 2011, 67(2): 513-523.

[10] Bender R, Augustin T, Blettner M. Generating survival times to simulate Cox proportional hazards models[J]. Statistics in Medicine, 2005, 24(11): 1 713-1 723.

[11] Wu H M. Kernel sliced inverse regression with applications to classification [J]. Journal of Computational and Graphical Statistics, 2008, 17(3): 590-610.

[12] Aronszajn N. Theory of reproducing kernels[J]. Transactions of the American Mathematical Society, 1950, 68(3): 337-404.

[13] Schölkopf B, Smola A J. Learning With Kernels: Support Vector Machines, Regularization, Optimization,and Beyond[M]. Cambridge, MA: MIT Press, 2001.

[14] Zhong W, Zeng P, Ma P, et al. RSIR: Regularized sliced inverse regression for motif discovery[J]. Bioinformatics, 2005, 21(22):4 169-4 175.

[15] Ferr W L, Villa N. Multilayer perceptron with functional inputs: An inverse regression approach[J]. Scandinavian Journal of Statistics, 2006, 33(4): 807-823.

[16] Li L, Yin X. Sliced inverse regression with regularizations[J]. Biometrics, 2008, 64(1):124-131.

[17] Cox D R. Regression models and life-tables[C]// Breakthroughs in Statistics. New York: Springer, 1992: 527-541.