带乘性噪声的各向异性回归模型小波估计 MISE

陈 佳

（四川文理学院数学学院，四川 达州 635000）

摘要：针对带乘法噪声回归估计模型，利用小波方法构造了回归函数线性小波估计器，并在回归函数具有各向异性和样本强混的条件下，在各向异性贝索夫空间中讨论了估计器的积分均方误差收敛阶。结果表明，当样本容量足够大时，小波估计器是相合的，且当空间的各向异性指标相等时，结论与各向同性贝索夫空间中的相关结果吻合。

关键词：乘性噪声、各向异性、小波估计、强混样本、积分均方误差。

一、引言

设 $\{(X_i,Y_i), i=1,2,\cdots,n\}$ 是概率空间 (Ω,F,P) 上的同分布随机变量，其公共密度函数为：

$$f(x,y)=\frac{\omega(x,y)g(x,y)}{\mu},(x,y)\in[0,1]^d\times\mathrm{R}, \quad (1)$$

其中 $\omega(x,y)$ 为偏差函数，$g(x,y)$ 表示未被观测到的随机变量 (U,V) 的密度函数，$\mu:=E(\omega(U,V))<\infty$（$EX$ 表示 X 的期望）。回归函数定义为：

$$r(x)=E(\rho(V)\,|\,U=x)=\frac{\int_{\mathrm{R}}\rho(y)g(x,y)dy}{h(x)},x\in[0,1]^d, \quad (2)$$

其中 ρ 为已知函数，h 为随机变量 U 的边际密度函数（即 $h(x)=\int_{\mathrm{R}}g(x,y)dy$）。

针对上述回归模型的研究可以追溯到 1964 年 Nadaraya 和 Watson 所做的工作，后来经 Chesneau、Chaubey 等人的研究，在样本独立的情形下讨论了估计器在各向同性贝索夫空间中的积分误差，文献给出了 $L^p(1\le p<\infty)$ 风险意义下小波估计的最优收敛速度；2017 年寇俊克针对带偏差非参数回归估计模型，考虑了多元小波估计量的 $L^p(1\le p<\infty)$ 风险收敛阶；2019 年，郭慧君基于独立随机样本在各向同性贝索夫空间中研究了回归函数的线性和非线性小波估计器的点态收敛阶。对于强混样本的情形，2018 年寇俊克用一种新的方法讨论了回归函数的 $L^p(1\le p<\infty)$ 风险的小波估计；2021 年，寇俊克给出了回归函数导数的线性和非线性小波估计器，并在 $B^s_{p,q}([0,1])$ 中研究了估计器的 L^2 风险收敛阶；2022 年，陈佳在各向同性贝索夫空间中研究了带乘法噪声回归函数的线性和非线性小波估计器的 $l^p(1\le p<\infty)$ 点态风险收敛阶；2022 年，陈佳针对强混带偏差样本回归模型，在各向异性贝索夫空间中讨论了线性小波估计器的点态 $l^p(1\le p<\infty)$ 风险。基于此，本文在各向异性贝索夫空间中研究强混样本下带乘性噪声回归函数的线性小波估计器的积分均分误差 (MISE)。

二、各向异性小波理论和贝索夫空间

设 $\{V_j: j\in\mathrm{Z}\}$ 是 $L^2(\mathrm{R})$ 中的多分辨率分析，$\tilde{\phi}$ 和 $\tilde{\psi}$ 为相应的 Daubechies 尺度函数和小波函数，因此 $\tilde{\phi}$，$\tilde{\psi}$ 满足（θ）条件（即 $\underset{x\in\mathrm{R}}{ess\sup}\sum_{k\in\mathrm{Z}}|h(x-k)|<+\infty$，$h$ 表示 $\tilde{\phi}$ 或 $\tilde{\psi}$）。

记 $\tau=(\tau_1,\cdots,\tau_d)$ 为 R^d 上的各向异性指标，满足 $\tau_i>0(1\le i\le d)$，$\sum_{i=1}^{d}\tau_i=d$，对 $\forall t>0$，$c\in\mathrm{R}$，$x\in\mathrm{R}^d$，约定

$$t^{\tau}x=(t^{\tau_1}x_1,\cdots,t^{\tau_d}x_d);\ t^{\tau c}=(t^c)^{\tau}.$$

定义指标集

$$\Omega_{j\tau}:=\left\{(\gamma,m):\gamma\in\{0,1\}^d\setminus\{0\}^d,m\in\mathrm{N}^d,j\in\mathrm{N}\right\},$$

且满足对任意的 $i=1,\cdots,d$，当 $\gamma_i=0$ 时，$m_i=\lfloor j\tau_i\rfloor$；当 $\gamma_i=1$ 时，$\lfloor j\tau_i\rfloor\le m_i<\lfloor(j+1)\tau_i\rfloor$。这里及以后，$\lfloor a\rfloor$ 表示 a 的整数部分，N^d 为分量均为自然数的 d 维全体向量。注意到集合 $\Omega_{j\tau}$ 的基数是有限的，事实上：

$$|\Omega_{j\tau}|\le(2^d-1)\prod_{i=1}^{d}((j+1)\tau_i+1-j\tau_i)\le(2^d-1)(d+1)^d<\infty.$$

定义 $\Phi_{j\tau;k}(x):=\prod_{i=1}^{d}2^{\frac{\lfloor j\tau_i\rfloor}{2}}\tilde{\phi}(2^{\lfloor j\tau_i\rfloor}x_i-k_i)$；$\Psi^{(\gamma,m)}_{j\tau;k}(x):=2^{\frac{|m|}{2}}\prod_{i=1}^{d}\psi^{\gamma_i}(2^{\lfloor m_i\rfloor}x_i-k_i)$, (3)

其中 $(\gamma,m)\in\Omega_{j\tau}$，$|m|=m_1+\cdots+m_d$，$\psi^{\gamma_i}=\begin{cases}\tilde{\phi},\ \gamma_i=0\\ \tilde{\psi},\ \gamma_i=1\end{cases}$。对任意的 $j_0\in\mathrm{N}$，

$$\left\{\Phi_{j_0\tau;k}(x),\Psi^{(\gamma,m)}_{j\tau;k}(x): j_0\le j\in\mathrm{N},(\gamma,m)\in I_{j\tau},k\in\mathrm{Z}^d\right\}, \quad (4)$$

构成 $L^2(\mathrm{R}^d)$ 的一组标准各向异性小波正交基，即对任意的 $f\in L^2(\mathrm{R}^d)$，

$$f(x)=\sum_{k\in\mathrm{Z}^d}\alpha_{j_0\tau;k}\Phi_{j_0\tau;k}(x)+\sum_{j=j_0}^{\infty}\sum_{(\gamma,m)\in I_{j\tau}}\sum_{k\in\mathrm{Z}^d}\beta^{(\gamma,m)}_{j\tau;k}\Psi^{(\gamma,m)}_{j\tau;k}(x), \quad (5)$$

其中 $\alpha_{j_0\tau;k}=\langle f,\Phi_{j_0\tau;k}\rangle$，$\beta^{(\gamma,m)}_{j\tau;k}=\langle f,\Psi^{(\gamma,m)}_{j\tau;k}\rangle$，在本文中 $\langle f,g\rangle=\int f(x)g(x)dx$。由于 $\tilde{\phi}$ 满足（θ）条件，可得 $\tilde{\phi}\in L(\mathrm{R})\cap L^{\infty}(\mathrm{R})$，则对于 $f\in L^p(\mathrm{R}^d)(1\le p<\infty)$，以及 $\mathrm{supp}f\subseteq[0,1]^d$，

$$f(x)=\sum_{k\in\Lambda_{j_0\tau}}\alpha_{j_0\tau;k}\Phi_{j_0\tau;k}(x)+\sum_{j=j_0}^{\infty}\sum_{(\gamma,m)\in I_{j\tau}}\sum_{k\in\Lambda_{j\tau}}\beta^{(\gamma,m)}_{j\tau;k}\Psi^{(\gamma,m)}_{j\tau;k}(x),$$

其中 $\Lambda_{j_0\tau}=\otimes_{i=1}^{d}\{1-2N,2-2N,\cdots,2^{\lfloor j_0\tau_i\rfloor}\}$，$\Lambda_{j\tau}=\otimes_{i=1}^{d}\{-N,1-N,\cdots,2^{\lfloor j\tau_i\rfloor}+N-1\}$。由 $\tilde{\phi}$ 和 $\tilde{\psi}$ 的紧支撑性可得集合 $\Lambda_{j_0\tau}$ 和 $\Lambda_{j\tau}$ 的基数有限，且 $|\Lambda_{j_0\tau}|\sim2^{|j_0\tau|}$，$|\Lambda_{j\tau}|\sim2^{|j\tau|}$。这里及以后，$A\prec B$ 表示 $A\le cB$，其中 c 为独立于 A 和 B 的正常数；$A\succ B$ 意指 $B\prec A$；$A\sim B$ 表示 $A\prec B$ 和 $B\prec A$。接下来，给出各向异性贝索夫空间定义。

定义 1 设 $k=(k_1,\cdots,k_d)\in\mathrm{Z}^d$，$\vec{s}=(s_1,\cdots,s_d)\in\mathrm{R}^d_+$，$k_i>s_i,i=1,\cdots,d$，对所有 $p,q\in[1,+\infty)$，$f\in B^{\vec{s}}_{p,q}(\mathrm{R}^d)$，当且仅当

(i) $f\in L^p(\mathrm{R}^d)$，(ii) $\|f\|_{B^{\vec{s}}_{p,q}}:=\|f\|_p+\sum_{i=1}^{d}\left(\int_0^1 t^{-s_iq}\left\|\Delta^{k_i}_{t_i}f(x)\right\|_p^q\frac{dt_i}{t_i}\right)^{\frac{1}{q}}<\infty$，

称 $B^{\vec{s}}_{p,q}(\mathrm{R}^d)$ 为各向异性贝索夫空间，是一个巴拿赫空间。

众所周知贝索夫空间可以用小波刻画，更多关于各向异性贝索夫空间的内容参考文献。

引理 1 设 $\frac{1}{s(d)}=\frac{1}{d}\sum_{i=1}^{d}\frac{1}{s_i}$，$s_i>0$，$\tau_i=\frac{s(d)}{s_i}(i=1,\cdots,d)$ 则 $f\in B^{\vec{s}}_{p,q}(\mathrm{R}^d)$ 等价于

$$\|f\|_{B^{s}_{p,q}}=\|\alpha_{j_0\tau.}\|_{l^p}+\left\{\sum_{j=j_0}^{\infty}\sum_{(\gamma,m)\in I_{j\tau}}\left[2^{j\left(s(d)-\frac{d}{p}\right)+\frac{|m|}{2}}\left\|\beta_{j\tau.}^{(\gamma,m)}\right\|_{l^p}\right]^q\right\}^{\frac{1}{q}}<+\infty .$$

其中 $\|\alpha_{j_0\tau.}\|_{l^p}^p=\sum_{k\in Z^d}|\alpha_{j_0\tau;k}|^p$，$\left\|\beta_{j\tau.}^{(\gamma,m)}\right\|_{l^p}^p=\sum_{k\in Z^d}\left|\beta_{j\tau;k}^{(\gamma,m)}\right|^p$。

备注 1：此定理蕴含着

$$\left\|\beta_{j\tau.}^{(\gamma,m)}\right\|_{l^p}\prec 2^{-j\left(s(d)-\frac{d}{p}\right)-\frac{|m|}{2}} \quad (6)$$

本文中还假设回归函数 $r(x)$ 属于 $H>0$ 的贝索夫球，即 $r(x)\in B^{\vec{s}}_{p,q}(H):=\left\{f\in B^{\vec{s}}_{p,q}(\mathrm{R}^d):\|f\|_{B^{\vec{s}}_{p,q}}\le H\right\}$。

三、线性小波估计量

在本节中，将对模型－给出一些假设，并构造线性小波估计器。

定义 2 设 $\{X_i, i\in Z\}$ 是严平稳的随机向量序列

$$\lim_{k\to\infty}\alpha(k)=\lim_{k\to\infty}\sup\left\{\left|P(A\cap B)-P(A)P(B)\right|:A\in F_{-\infty}^0, B\in F_k^{\infty}\right\}=0,$$

则序列 $\{X_i, i\in Z\}$ 是强混合的。其中 $F_{-\infty}^0$ 表示 $\{X_i, i\le 0\}$ 所生成的 σ －代数，F_k^{∞} 由 $\{X_i, i\ge k\}$ 所生成的 σ －代数。下面，对模型－给出一些假设：

H1、存在常数 $c_1>0$，使得 $\inf_{x\in[0,1]^d}h(x)\ge c_1$。

H2、ρ 满足 $\rho\in L(R^d)\cap L^{\infty}(R^d)$。

H3、对任意 $(x,y)\in[0,1]^d\times\mathrm{R}$，存在常数 $0<c_2<c_3<\infty$ 使得 $0<c_2<\omega(x,y)<c_3<\infty$。

H4、序列 $\{(X_i,Y_i), i=1,\cdots,n\}$ 的强混系数满足 $\alpha(k)\le\lambda e^{-c_4k}$，其中 $\lambda>0$，$c_4>0$。

H5、设 $f_{(X_1,Y_1,X_{k+1},Y_{k+1})}$ 为 $(X_1,Y_1,X_{k+1},Y_{k+1})(k\ge1)$ 的密度函数，$f_{(X_1,Y_1)}$ 为 (X_1,Y_1) 的密度函数，存在常数 $c_5>0$，使得 $\forall(x,y,x',y')\in[0,1]^d\times\mathrm{R}\times[0,1]^d\times\mathrm{R}$，

$$\sup_{k\ge1}\sup_{(x,y,x',y')\in[0,1]^d\times\mathrm{R}\times[0,1]^d\times\mathrm{R}}\left|h_k(x,y,x',y')\right|\le c_5,$$

其中 $h_k(x,y,x',y')=f_{(X_1,Y_1,X_{k+1},Y_{k+1})}(x,y,x',y')-f_{(X_1,Y_1)}(x,y)f_{(X_1,Y_1)}(x',y')$。

各向异性回归函数 $r(x)$ 的线性小波估计量如下，

$$\hat{r}_{j\tau}(x)=\sum_{k\in\Lambda_{j\tau}}\hat{\alpha}_{j\tau;k}\Phi_{j\tau;k}(x) \quad (7)$$

其中

$$\hat{\alpha}_{j\tau;k}=\frac{\hat{\mu}_n}{n}\sum_{i=1}^{n}\frac{\rho(X_i)}{\omega(X_i,Y_i)h(X_i)}\Phi_{j\tau;k}(X_i), \quad (8)$$

$$\hat{\mu}_n=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right]^{-1} \quad (9)$$

此估计量是 $r(x)$ 的一个无偏估计，从以下的引理2可得到证实。

引理 2 考虑由－定义的问题，$\hat{\mu}_n$ 由定义，那么

$$E\left(\hat{\mu}_n^{-1}\right)=\mu^{-1},\quad E\left(\frac{\mu\rho(X_i)}{\omega(X_i,Y_i)h(X_i)}\Phi_{j\tau;k}(X_i)\right)=\alpha_{j\tau;k} .$$

证明：根据期望和 $\hat{\mu}_n$ 的定义，且 $(X_1,Y_1),\cdots,(X_n,Y_n)$ 同分布，

$$E\left(\hat{\mu}_n^{-1}\right)=E\left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\omega(X_i,Y_i)}\right]=E\left(\frac{1}{\omega(X_1,Y_1)}\right)=\int_{[0,1]^d\times\mathrm{R}}\frac{1}{\omega(x,y)}f(x,y)dxdy=\mu^{-1} .$$

另一方面

$$E\left(\frac{\mu\rho(X_i)}{\omega(X_i,Y_i)h(X_i)}\Phi_{j\tau;k}(X_i)\right)=\int_{[0,1]^d\times\mathrm{R}}\frac{\mu\rho(x)}{\omega(x,y)h(x)}\Phi_{j\tau;k}(x)f(x,y)dxdy=\alpha_{j\tau;k} .$$

为得到本文的结论，还需要引用以下引理。

引理 3 设 $\hat{\alpha}_{j\tau;k}$ 由 (8) 定义，假设 H1–H5 成立且 $2^{|j\tau|}\le n$，则

$$E\left|\hat{\alpha}_{j\tau;k}-\alpha_{j\tau;k}\right|^2\prec n^{-1} .$$

四、本文结论

本节将给出本文主要结果及其相应的证明。

定理 设 $r\in B^{\vec{s}}_{p,q}(H)\left(p,q\in[1,\infty)\right)$，$\hat{r}_{j\tau}(x)$ 由式定义，假设 H1–H5 成立且 $2^{\lfloor j\tau_i\rfloor}\sim n^{\frac{s(d)}{2(s(d)-d/p)+d}\frac{1}{s_i}}$，则

$$E\int_{[0,1]^d}\left|\hat{r}_{j\tau}(x)-r(x)\right|^2dx\prec n^{-\frac{2(s(d)-d/p)}{2(s(d)-d/p)+d}} .$$

证明：根据 $\tilde{\phi}$，$\tilde{\psi}$ 的正交性，易知

$$E\int_{[0,1]^d}\left|\hat{r}_{j\tau}(x)-r(x)\right|^2dx=E\left\|r_{j\tau}(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2+\left\|r(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2 .(10)$$

因此，只需要估计式右侧的随机项和偏差项。

先估计随机项，由于 $\Phi_{j\tau;k}$ 具有正交性，所以

$$E\left\|\hat{r}_{j\tau}(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2=E\sum_{k\in\Lambda_{j\tau}}\left|\hat{\alpha}_{j\tau;k}-\alpha_{j\tau;k}\right|^2\cdot\left\|\Phi_{j\tau;k}(x)\right\|_2^2$$

根据引理 3 以及 $\left\|\Phi_{j\tau;k}(x)\right\|_2=1$，$2^{\lfloor j\tau_i\rfloor}\sim n^{\frac{s(d)}{2(s(d)-d/p)+d}\frac{1}{s_i}}$，则

$$E\left\|\hat{r}_{j\tau}(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2=\sum_{k\in\Lambda_{j\tau}}E\left|\hat{\alpha}_{j\tau;k}-\alpha_{j\tau;k}\right|^2\prec n^{-1}2^{|j\tau|}\prec n^{-1}n^{\frac{s(d)}{2(s(d)-d/p)+d}\sum_{i=1}^{d}\frac{1}{s_i}}=n^{-\frac{2(s(d)-d/p)}{2(s(d)-d/p)+d}} \quad (11)$$

再估计偏差项，利用赫尔德不等式可得，

$$\left\|r(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2\le\int_{[0,1]^d}\left|\sum_{l=j}^{\infty}\sum_{(\gamma,m)\in\Omega_{l\tau}}\left(\sum_{k\in\Delta_{l\tau}}\left|\beta_{l\tau;k}^{(\gamma,m)}\right|^p\left|\Psi_{l\tau;k}^{(\gamma,m)}(x)\right|\right)^{\frac{1}{p}}\left(\sum_{k\in\Delta_{l\tau}}\left|\Psi_{l\tau;k}^{(\gamma,m)}(x)\right|\right)^{\frac{1}{p'}}\right|^2dx\left(\frac{1}{p}+\frac{1}{p'}=1\right)$$

根据 $\Phi_{j\tau;k}$ 的紧支性和（ θ ）条件进一步可得

$$\left\|r(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2\prec\int_{[0,1]^d}\left|\sum_{l=j}^{\infty}\sum_{(\gamma,m)\in\Omega_{l\tau}}2^{\frac{|m|}{2}}\left(\sum_{k\in\Delta_{l\tau}}\left|\beta_{l\tau;k}^{(\gamma,m)}\right|^p\right)^{\frac{1}{p}}\right|^2dx\prec\left(\sum_{l=j}^{\infty}\sum_{(\gamma,m)\in\Omega_{l\tau}}2^{\frac{|m|}{2}}2^{-l(s(d)-d/p)-\frac{|m|}{2}}\right)^2\prec 2^{-2j(s(d)-d/p)} .$$

根据 $2^{\lfloor j\tau_i\rfloor}\sim n^{\frac{s(d)}{2(s(d)-d/p)+d}\frac{1}{s_i}}$ 和 $|j\tau|=\sum_{i=1}^{d}\lfloor j\tau_i\rfloor\le\sum_{i=1}^{d}j\tau_i=jd$，上式可化为

$$\left\|r(x)-\sum_{k\in\Lambda_{j\tau}}\alpha_{j\tau;k}\Phi_{j\tau;k}(x)\right\|_2^2\prec 2^{-2j(s(d)-d/p)}\le 2^{-\frac{2}{d}|j\tau|(s(d)-d/p)}\prec\left(n^{\frac{s(d)}{2(s(d)-d/p)+d}\sum_{i=1}^{d}\frac{1}{s_i}}\right)^{-\frac{2}{d}(s(d)-d/p)}\prec n^{-\frac{2(s(d)-d/p)}{2(s(d)-d/p)+d}} \quad (12)$$

综上所述，结合定理得证。

五、结束语

本文利用小波方法在各向异性贝索夫空间中间构造了回归函数的线性小波估计器 $\hat{r}(x)$，并讨论估计器在了 L^2 风险意义下的收敛阶 (即 MISE)。结论表明，当样本容量足够时，估计器依积分均分误差 (MISE) 的收敛速阶不超过 $n^{-\frac{2(s(d)-d/p)}{2(s(d)-d/p)+d}}$。

参考文献：

[1] 郭慧君，寇俊克 . 基于偏置数据的回归函数点态小波估计 [J]. 数学结果，2019，74（4）：1–14.

[2] 寇俊克，刘有明 . 强混数据的小波回归估计 [J], 统计学方法及应用，2018，27（4）：667 – 688.

[3] 寇俊克，陈佳，郭慧君 . 基于偏置强混数据的回归函数导数的小波估计 [J]. 统计通讯：理论与方法，2021，50（14）：3436–3452.

基金项目：四川文理学院校级科研项目（2022QD06）

作者简介：陈佳（1992—），男，四川南充，硕士研究生，助教，主要从事小波分析及其应用与非参数估计研究。