应用概率统计　第 37 卷
第 2 期　2021 年 4 月

Chinese Journal of Applied Probability and Statistics
Apr., 2021, Vol. 37, No. 2, pp. 201-220
doi: 10.3969/j.issn.1001-4268.2021.02.008

综述报告

小波估计方法发展综述 *

邹玉叶 [1,2★]　　范国良 [2]

([1] 华东师范大学统计学院和统计与数据科学前沿理论及应用教育部重点实验室, 上海, 200062)

([2] 上海海事大学经济管理学院, 上海, 201306)

摘　要： 小波估计方法一直是统计学领域中的研究热点和难点问题, 在数据压缩、流体湍流、信号和图像处理、地震勘探等领域有着广泛的应用价值. 本文以小波估计方法在数理统计中的应用为研究对象, 重点介绍小波估计方法的基本理论、门限函数种类, 以及小波估计方法在完全数据、不完全数据和纵向数据下的研究成果. 由于数据的复杂性和不完全性, 导致传统的研究方法不再适用, 需要结合左截断数据、右删失数据、缺失数据和纵向数据的特点, 利用插入法、回归校正法、插补法和可逆概率加权法, 构造被估函数的非线性小波估计量, 研究非线性小波估计量平均积分二次误差 (mean integral square error, MISE) 的渐近展开式和估计量的渐近正态性; 讨论被估函数存在有限个不连续点时, 非线性小波估计量 MISE 仍然成立; 证明非线性小波估计量在包含很多不连续函数的 Besov 空间里的一致收敛性; 利用小波估计方法研究回归模型中参数和非参数估计量的相合性和收敛速度; 最后简要探讨小波估计方法未来的可能发展方向.

关键词： 渐近性质; 不完全数据; 线性小波估计; 纵向数据; 非线性小波估计

中图分类号： O212.7

英文引用格式： ZOU Y Y, FAN G L. Development review of wavelet estimation method [J]. Chinese J Appl Probab Statist, 2021, 37(2): 201–220. (in Chinese)

§1.　引　　言

自 1882 年 Fourier 发表热传导解析理论以来, 调和分析在数学、物理学以及工程学中得到广泛应用. 小波分析是 1986 年以来在 Mallat [1]、Daubechies [2] 及 Daubechies 和 Heil [3] 等奠基工作的基础上迅速发展起来的一门应用数学学科, 它是傅里叶分析发展史上的里程碑, 也是调和分析的工作结晶. 原则地讲, 传统上使用傅里叶分析的地方, 现在都可以用小波分析取代. 小波的理论成果很丰富, 实际应用也很广泛 [4]. 例如在机械工程学中, 利用小波理论进行信息和图像处理、去噪 [5,6]; 在地质学中, 利用小波理论进行地震勘探.

*中国博士后科学基金项目 (批准号: 2019M651422) 资助.

★通讯作者, E-mail: zouyuye@shmtu.edu.cn.

本文 2019 年 8 月 21 日收到, 2019 年 10 月 8 日收到修改稿.

小波变换被称作 “数学显微镜”, 能够提高非线性、非平稳性信号的时频分辨率, 具有很广泛的应用前景. 另外, 小波理论在理论数学中的发展也很乐观, 常被用来研究不规则函数的渐近性质、变点检测、函数型数据降维等. 本文将主要介绍小波理论在数理统计中的发展和研究成果, 最后给出一些公开问题, 并且指出新的研究方向.

用 $L_p(\mathbb{R})$ $(1 \leqslant p < \infty)$ 表示 p 方可积的函数空间, 即

$$L_p(\mathbb{R}) = \Big\{ f : \int_{-\infty}^{\infty} |f(x)|^p \mathrm{d}x < \infty \Big\},$$

其范数定义为

$$\|f\|_p = \Big[\int_{-\infty}^{\infty} |f(x)|^p \mathrm{d}x \Big]^{1/p},$$

当 $p = 2$ 时, 平方可积函数空间 $L_2(\mathbb{R})$ 即为 Hilbert 空间.

长期以来, 学者们一直努力寻找 $L_2(\mathbb{R})$ 空间里的一种既可以保持指数基优点, 又可以弥补指数基不足的函数基; 并且这种函数基是由某个具有光滑性、紧支撑性和较高的消失矩的函数通过伸缩和平移而生成的函数族. 现在我们称这种函数基为 “小波基”. 第一个真正的小波基是由 1986 年 Meyer 在怀疑小波基的存在性时构造出来的. 在给出相关理论成果之前, 首先介绍小波基的定义、性质和一些记号. 令父小波函数 $\varphi(t)$ 和母小波函数 $\psi(t)$ 有界和紧支撑, 并且 $\int \varphi^2 = \int \psi^2 = 1$, $\mu_k = \int t^k \psi(t) \mathrm{d}t = 0$ 对于 $0 \leqslant k \leqslant r-1$, $\mu_r = r!\kappa \neq 0$, 其中 $\kappa = (r!)^{-1} \int t^r \psi(t) \mathrm{d}t$, r 是正整数. 令

$$\varphi_{k_0 l}(t) = 2^{k_0/2} \varphi(2^{k_0} t - l), \quad \psi_{kl}(t) = 2^{k/2} \psi(2^k t - l), \qquad t \in \mathbb{R},$$

对于任意 k_0, $k \geqslant k_0$, $l \in \mathbb{Z}$. 于是 $\{\varphi_{k_0 l}(t), \psi_{kl}(t), k, l \in \mathbb{Z}, k \geqslant k_0\}$ 形成 $L_2(\mathbb{R})$ 空间的一组正交基. 正则关系表现如下,

$$\int \varphi_{k_0 l_1} \varphi_{k_0 l_2} = \delta_{k_0 l_1} \delta_{k_0 l_2}, \qquad \int \psi_{k_1 l_1} \psi_{k_2 l_2} = \delta_{k_1 l_1} \delta_{k_2 l_2}, \qquad \int \varphi_{l_1} \psi_{k_2 l_2} = 0,$$

其中 δ_{ij} 表示克罗内克系数, 即 $\delta_{ij} = 1$, 如果 $i = j$; 否则为 0. 更多关于小波记号可以参见文献 [7].

对任意的函数 $f(x) \in L_2(\mathbb{R})$ 可以分解成如下小波形式:

$$f(x) = \sum_l b_{k_0 l} \varphi_{k_0 l}(x) + \sum_{k=k_0}^{\infty} \sum_l b_{kl} \psi_{kl}(x), \tag{1}$$

其中, $b_{k_0 l} = \int \varphi_{k_0 l}(x) f(x) \mathrm{d}x$ 和 $b_{kl} = \int \psi_{kl}(x) f(x) \mathrm{d}x$ 是函数 $f(x)$ 的小波系数. 展开式 (1) 是一种特殊的正交级数. 它的特殊性主要体现在它不像普通的傅里叶级数, 它的近似性是在频率和空间上体现. 关于小波估计方法的研究可参见文献 [8, 9].

构造 $f(x)$ 估计量的基本思想是基于观测数据, 分别用小波系数的估计量取代 (1) 式中的小波系数 $b_{k_0 l}$ 和 b_{kl}. 这就需要给 (1) 式中的无穷级数一个截断, 因为我们只能够处理有

限个小波系数. 根据表现形式不同, 小波估计可以分成两种: 线性小波估计和非线性小波估计. 非线性小波估计量的构造是依赖于门限, 根据门限形式不同, 非线性小波估计量可以分为局部门限非线性小波估计量, 全局门限非线性小波估计量和块门限非线性小波估计量 (详见第 3 部分第 1 小节).

本文首先介绍小波估计方法的基本知识. 第 2 部分主要介绍完全数据下和纵向数据下被估函数的线性小波估计. 第 3 至第 8 部分主要介绍完全数据、右删失数据、左截断数据、缺失数据下非线性小波估计量的构造和研究成果. 第 9 部分我们对小波估计方法的研究成果进行总结, 并对小波估计方法未来的理论研究和实际应用进行展望.

§2.　线性小波估计

1) 完全数据线性小波估计

Donoho 等[19] 构造了随机变量序列 $\{X_i,\ 1 \leqslant i \leqslant n\}$ 密度函数 $f(x)$ 的线性小波估计量

$$\widehat{f}_n(x)=\sum_l \widehat{b}_{k_0l}\varphi_{k_0l}(x)+\sum_{k=k_0}^{k_1-1}\sum_l \widehat{b}_{kl}\psi_{kl}(x), \tag{2}$$

其中 $k_0, k_1 \in \mathbb{Z}$ 是光滑参数,

$$\widehat{b}_{k_0l}=\frac{1}{n}\sum_{i=1}^{n}\varphi_{k_0l}(X_i), \qquad \widehat{b}_{kl}=\frac{1}{n}\sum_{i=1}^{n}\psi_{kl}(X_i)$$

是小波系数的无偏经验估计量. 假设父小波 φ 和母小波 ψ 有紧支撑, 注意到对于任意的

$$\sum_l \varphi_{kl}(X_i)\varphi_{kl}(x)+\sum_l \psi_{kl}(X_i)\psi_{kl}(x)=\sum_l \varphi_{k+1,l}(X_i)\varphi_{k+1,l}(x)=K_{k+1}(x,X_i),$$

其中正交投影核定义如下,

$$K_k(x,y)=2^kK(2^kx,2^ky), \qquad K(x,y)=\sum_l \varphi(x-l)\varphi(y-l).$$

于是, 定义如下形式的线性小波估计量

$$\widehat{f}_n(x)=\sum_l \widehat{b}_{k_1+1,l}\varphi_{k_1+1,l}(x)=\frac{1}{n}\sum_{i=0}^{n}K_{k+1}(x,X_i), \tag{3}$$

这里的 k 是从 k_0 开始, 与 (1) 式中 $k=0$ 开始不同. 这与一般理论不冲突, 假设 $k_0=0$ 是为了简化记号.

Besov 空间 $\{B_{pq}^s,\ s>0, 1\leqslant p,q\leqslant\infty\}$, 其中 s 是光滑参数, p 和 q 用于指定范数的类型. Besov 空间包括很多传统的函数空间, 特别是光滑函数 H^m 和 C^m 的 Sobolev 空间及

Hölder 空间 (分别是 B_{22}^m 和 $B_{\infty\infty}^s$). 另外, 还包括具有显著空间不均匀性的函数类, 例如块代数函数类和有界变量函数类. 根据小波系数, 如果 $f\in B_{pq}^s$, 当且仅当

$$\|f\|_{B_{pq}^s}=\|b_{k_0l}\|+\Big(\sum_{k\geqslant k_0}2^{kq(s+1/2-1/p)}\|b_{kl}\|_p^q\Big)^{1/q}<\infty,$$

其中 $\|b_{k_0l}\|_p=\big(\sum\limits_l|b_{k_0l}|^p\big)^{1/p}$ 和 $\|b_{kl}\|_p=\big(\sum\limits_l|b_{kl}|^p\big)^{1/p}$. 与 Sobolev 空间相比, Besov 空间的优点是它们在描述函数的光滑性方面更通用. Besov 空间允许用小波系数来表征, 而 Sobolev 空间则不允许. Besov 空间与小波分析曲线有着内在的联系.

Donoho 等[10] 给出密度函数 $f(\cdot)$ 在包含很多不连续函数类的 Besov 空间里线性小波估计量的次最优收敛速度. Kerkyacharian 和 Picard[11]及 Walter[12] 分别讨论了线性小波估计量的收敛速度.

序列 $\{\xi_i,\,i\geqslant1\}$ 被称为 α 混合数列, 如果 α 混合数列

$$\alpha(m):=\sup_{k\geqslant1}\sup\{|P(AB)-P(A)P(B)|:A\in\mathfrak{F}_{m+k}^{\infty},\,B\in\mathfrak{F}_1^k\}$$

收敛到 0 当 $m\to\infty$, 其中 $\mathfrak{F}_a^b=\sigma\{\xi_i,\,a\leqslant i\leqslant b\}$ 表示由 $\xi_a,\xi_{a+1},\cdots,\xi_b$ 生成的 σ 代数. 在已有文献使用的各种混合条件中, α 混合较弱, 在包含一些时间序列模型的很多随机过程中使用广泛. 实际上, 在温和的假设条件下, 线性自回归和更一般的双线性时间序列模型混合程度强, 当混合系数退化成指数, 即 $\alpha(k)=O(\rho^k)$ 对于一些 $0<\rho<1$, 详见文献 [13].

Liang[14] 讨论了相依样本异方差模型小波估计量的渐近性质. Leblanc[15] 将 Kerkyacharian 和 Picard[11] 的结果推广到相依样本, 建立了离散时间线性小波估计量 L_p 误差的界. Masry[16] 研究了相依样本概率密度函数线性小波估计量 MISE 的渐近性质. 李永明和韦程东[17] 证明了相依样本回归函数小波估计量的 Berry-Essen 的界.

2) 半参数回归模型线性小波估计

半参数回归模型在国内外具有广泛的基础研究. 小波估计方法在半参数回归模型里的应用也有很多学者关注. 考虑如下形式的半参数回归模型

$$y_i=x_i\beta+g(t_i)+\epsilon_i,\qquad i=1,2,\cdots,n,\tag{4}$$

其中, x_i, t_i 是协变量, y_i 是反应变量, β 是未知参数, $g(\cdot)$ 是连接函数, ϵ_i 是随机误差.

通过父小波函数 $\varphi(\cdot)$, 定义如下形式的小波核

$$E_m(t,s)=2^mE_0(2^mt,2^ms)=2^m\sum_{k\in\mathbb{Z}}\varphi(2^mt-k)\varphi(2^ms-k).$$

令 $s_n^2=\sum\limits_{i=1}^n\widetilde{x}_i^2$,

$$\widetilde{x}_i=x_i-\sum_{j=1}^nx_j\int_{A_j}E_m(t_i,s)\mathrm{d}s,\qquad \widetilde{y}_i=y_i-\sum_{j=1}^ny_j\int_{A_j}E_m(t_i,s)\mathrm{d}s,$$

其中 $A_j=[s_{j-1},s_j]$, $s_0=0$, $s_n=1$, $s_j=(t_j+t_{j+1})/2$, $j=1,2,\cdots,n-1$. 于是利用最小二乘法构造未知参数 β 的小波估计量

$$\widehat{\beta}_n=\sum_{j=1}^{n}\widetilde{x}_i\widetilde{y}_i/s_n^2.$$

再利用插入法, 构造非参数部分 $g(t_i)$ 的小波估计量

$$\widehat{g}_n(t_i)=\sum_{i=1}^{n}(y_i-x_i\widehat{\beta}_n)\int_{A_i}E_m(t_i,s)\mathrm{d}s.$$

薛留根[18] 把小波光滑和随机加权方法结合, 讨论参数 β 小波估计的误差分布随机加权逼近问题, 并证明它的逼近精度可以达到 $o(n^{-1/2})$. 胡宏昌和胡迪鹤[19] 讨论了半参数回归模型参数和函数小波估计量的强相合性. 刘强和薛留根[20] 利用最小二乘法和小波估计定义了参数和函数的估计量, 在误差序列为 ψ 混合或是 φ 混合情况下, 讨论了参数估计量的强相合性, 以及函数估计量的一致强相合性和 r 阶矩相合性. Xue 和 Liu[21] 采用 bootstrap 方法、小波估计和 Efron 再抽样技术构造了 bootstrap 统计量, 证明了 bootstrap 逼近的强一致收敛性, 并构造了参数的大样本置信区间. Liang 和 Wang[22] 在相依 MA(∞) 误差下研究了参数 β 和函数 $g(\cdot)$ 小波估计量的强一致收敛速度, 结果表明, $g(\cdot)$ 小波估计量可以达到最优收敛速度.

3) 纵向数据线性小波估计

纵向数据是指对同一组受试个体在不同时间和空间上的观测数据. 这类数据在生物医学、社会科学、计量经济学和流行病学领域有着广泛的应用性. 考虑有 n 个个体的样本, 对第 i 个个体在时间点 $t=t_{ij}$ $(i=1,2,\cdots,n,\ j=1,2,\cdots,m)$ 处对响应变量 $Y_i(t)$ 和协变量 $X_i(t)$ 进行观测, 其中 m_i 表示对第 i 个个体总的观测次数, 那么可得到纵向数据 $\{(t_{ij},X_i(t_{ij}),Y_i(t_{ij})),\ i=1,2,\cdots,n,\ j=1,2,\cdots,m\}$. 对纵向数据的研究可参见文献 [23–25] 等.

在纵向数据下, 关于小波估计方法也有学者研究. 刘刚等[26] 考虑了纵向数据下变系数回归模型

$$y_{ij}=x_{ij}^{\mathsf{T}}g(t_{ij})+e_{ij},\qquad i=1,2,\cdots,n, \tag{5}$$

其中 $(x_{ij}^{\mathsf{T}},t_{ij})\in\mathbb{R}^d\times p$ 是固定已知的设计点列, $g(\cdot)$ 是 d 维的未知函数向量, e_{ij} 是随机误差, 且 $\mathsf{E}(e_{ij})=0$, $\mathsf{Var}\,(e_{ij})=\sigma^2<\infty$. 为了方便定义估计量, 给出以下记号 $x_i=(x_{i1},x_{i2},\cdots,x_{im})^{\mathsf{T}}$, $y_i=(y_{i1},y_{i2},\cdots,y_{im})^{\mathsf{T}}$, $e_i=(e_{i1},e_{i2},\cdots,e_{im})^{\mathsf{T}}$, $X=(x_1,x_2,\cdots,x_n)^{\mathsf{T}}$, $Y=(y_1,y_2,\cdots,y_n)^{\mathsf{T}}$, $e=(e_1,e_2,\cdots,e_n)^{\mathsf{T}}$, $W(t)=\mathrm{diag}(\int_{A_1}E_m(t,s),\int_{A_2}E_m(t,s),\cdots,\int_{A_n}E_m(t,s))$, 则函数 $g(t)$ 的小波估计量定义为

$$\widehat{g}_n(t)=[X^{\mathsf{T}}W(t)X]^{-1}X^{\mathsf{T}}W(t)Y,$$

令 $\widehat{e}_{ij}=y_{ij}-x'_{ij}\widehat{g}_n(t_{ij})$, 于是定义误差方差的小波估计量

$$\widehat{\sigma}_n^2=\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\widehat{e}_{ij}^2.$$

在适当的假设条件下, 刘刚等[26] 证明未知函数 $g(t)$ 小波估计量的强相合性, 强相合收敛速度和渐近正态性, 同时给出了误差方差 σ^2 估计量的强相合性和渐近正态性.

研究结果表明, 线性小波估计量可以达到次最优收敛速度, 但是当被估计函数是非齐次的或是具有未知正则性时, 线性小波估计量不能达到最优收敛速度. 然而, 小波门限提供了一种自动适用被估计函数正则性的方法, 基于小波门限构造的非线性小波估计量可以达到最优收敛速度.

§3. 完全数据非线性小波估计

1) 非线性小波估计的门限函数

在完全独立数据下, 当未知密度函数 $f(\cdot)$ 在包含很多不连续函数类的 Besov 空间时, Donoho 等[10] 证明了线性小波估计量的最优收敛速度是 $O(n^{-2s'/(2s'+1)})$, 其中 $s'=s+1/2-1/p$. 注意到当 $s>s'$ 时, 有 $2s/(2s+1)>2s'/(2s'+1)$. 当 $p<2$ 时, $f(\cdot)$ 的非线性小波估计量的收敛速度比线性收敛速度要快一些. 主要是因为线性小波估计量可能有比较小的峰值, 包括不必要的高震荡, 这种高震荡是由小波系数 b_{kl} 产生的. 自然地, 需要介绍一种选择小波系数 b_{kl} 的方法, 更准确地说, 通过引进一个门限可以抑制较小的小波系数. 常用的小波门限主要分三种: 局部门限、全局门限和块门限. 下面给出这三种小波门限的定义式, 令函数 $\eta(u)$ 是依赖于随机变量 $\{X_i,\ 1\leqslant i\leqslant n\}$ 的随机函数, 假设

$$\eta(u)=0,\qquad |u|\leqslant t,$$

其中, $t>0$ 是一个门限 (可能随机). η_{kl} 是非随机的, 并且不依赖 k,l. 现定义局部门限, 局部门限分两种: 局部软门限和局部硬门限,

$$\text{局部软门限}:\eta_{kl}(u)=\eta^S(u)=(|u|-t)_+\operatorname{sign}u;$$
$$\text{局部硬门限}:\eta_{kl}(u)=\eta^H(u)=uI(|u|>t).$$

根据上述门限, 可以构造如下形式的局部门限非线性小波估计量

$$\widehat{f}_L(x)=\sum_l\widehat{b}_{k_0l}\varphi_{k_0l}(x)+\sum_{k=k_0}^{k_1-1}\sum_l\eta_{kl}(\widehat{b}_{kl})\psi_{kl}(x).\tag{6}$$

Kerkyacharian 等[27] 考虑了如下两种形式的全局门限

$$\text{全局软门限}:\eta_j^S(u)=uI\Big(\frac{S_j(p)-2^j/n^{p/2}}{S_j(p)}\Big)_+;$$

$$\text{全局硬门限}：\eta_j^H(u) = uI\Big(S_j(p) > \frac{2^j}{n^{p/2}}\Big),$$

其中 $S_j(p)$ 是依赖于 $\{X_1, X_2, \cdots, X_n\}$ 的确定统计量, $p \geqslant 1$ 是参数. 我们可以不保留或不删除单个小波系数, 也可以保留或删除整个 j 级系数, 于是可以构造如下形式的全局门限非线性小波估计量

$$\widehat{f}_G(x) = \sum_l \widehat{b}_{k_0 l}\varphi_{k_0 l}(x) + \sum_{k=k_0}^{k_1-1} \eta_j\Big[\sum_l \widehat{b}_{kl}\psi_{kl}(x)\Big]. \tag{7}$$

块门限介于局部门限和全局门限之间, Hall 等[28] 指出块门限在每一级都保留或删除特殊选定的小波系数块. 块门限把所有整数集分成长度为 $\{j = j(n)\}$ 的不重叠块

$$B_k = \{m : (k-1)j + 1 \leqslant m \leqslant kj\}, \qquad k \in \mathbb{Z}.$$

令 $\widehat{b}_{kl} = l^{-1} \sum\limits_{m \in B_k} \widehat{b}_{lm}$ 是 $b_{kl} = l^{-1} \sum\limits_{m \in B_k} b_{lm}$ 估计量. 于是可以定义如下形式的块门限非线性小波估计量

$$\widehat{f}_B(x) = \sum_l \widehat{b}_{k_0 l}\varphi_{k_0 l}(x) + \sum_{k=k_0}^{k_1} \sum_l \Big[\sum_{m \in B_k} \widehat{b}_{kl}(x)\psi_{km}(x)\Big] I(\widehat{b}_{kl} > Cn^{-1}), \tag{8}$$

其中 $C > 0$ 是控制门限的常数. 在很多情况下, 块门限非线性小波估计量比局部门限非线性小波估计量具有更好的渐近性质, 因为它具有不带额外的对数 ln 项的收敛速度.

2) 完全数据未知函数非线性小波估计

在完全数据下, 小波估计的理论成果有很多文献可以查询. Hall 和 Patil[29] 给出完全数据下核密度估计量的 MISE 的渐近展开式

$$\text{MISE} \sim C_1(nh)^{-1} + C_2 h^{2r},$$

$\sim$ 表示当样本容量 $n \to \infty$ 时上式符号左右两边的比值趋于 1, $0 < h \to 0$ 是核函数的窗宽, r 是核函数的阶, C_1 和 C_2 是依赖于核函数和密度函数的常数. 如果密度函数不存在 r 阶导数时, 那么密度函数核估计量 MISE 的渐近展开式不成立. 然而, 不管未知密度函数光滑与否, 密度函数的非线性小波估计量 MISE 的渐近展开式仍然成立.

Hall 和 Patil[29] 构造完全独立样本密度函数非线性小波估计量

$$\widehat{f}_n(x) = \sum_l \widehat{\alpha}_l\varphi_l(x) + \sum_{k=0}^{q-1} \sum_l \widehat{\alpha}_{kl} I(|\widehat{\alpha}_{kl}| > \lambda)\psi_{kl}(x), \tag{9}$$

其中 λ 是门限系数, q 是光滑参数, $\widehat{\alpha}_l$ 和 $\widehat{\alpha}_{kl}$ 是小波系数估计量. 他们首次论证了未知函数不连续性对非线性小波估计量的影响是可以忽略的, 给出了非线性小波估计量 MISE 的渐近展开式. Patil[30] 证明了完全独立样本风险率函数局部线性小波估计量的渐近性质.

在已有文献中, 绝大多数未知函数非线性小波估计量都是由经验小波局部门限构造的, 可以得到带有对数 ln 项的最优收敛速度. Hall 等[28,31] 定义了完全独立样本密度函数和回归函数块门限非线性小波估计量, 并得到不带对数 ln 项的最优收敛速度. Zhang 和 Zheng[32] 讨论了随机设计的回归函数非线性小波估计量的渐近性质. Cai[33] 通过改进块门限系数, 讨论了回归函数块门限非线性小波估计量渐近收敛速度. Kerkyacharian 和 Picard[34] 利用核估计方法和非线性小波估计法, 研究密度函数估计量在 Besov 空间的渐近性质. 薛留根[35] 证明了混合误差下回归函数小波估计量的一致收敛速度.

Li 和 Xiao[36] 考虑如下非参数回归模型

$$Y_k = g(x_k) + \epsilon_k, \qquad k = 1, 2, \cdots, n,$$

其中 $x_k = k/n \in [0,1]$, $g(\cdot)$ 是未知函数. 令 $\{\epsilon_k, k \geqslant 1\}$ 是均值为 0, 方差为常数的平稳过程, 如果有 $\sum\limits_{k=1}^{\infty} |\rho(k)| = \infty$, 其中 $\rho(k) = \mathsf{E}(\epsilon_k \epsilon_{k+1})$ 是 ϵ_k 的自回归协方差函数, 则称 $\{\epsilon_k\}$ 为长程相依. Li 和 Xiao[36] 基于完全长程相依样本, 考察了当平均回归函数仅点点光滑时, 平均回归函数的非线性小波估计量 MISE 渐近展开式, 与核估计量 MISE 展开式一样. 然而, 如果被估函数不具备光滑性的条件, 核估计量 MISE 展开式不成立.

§4. 右删失数据非线性小波估计

在生存分析中, 病人在进行某种疾病的治疗时, 由于某些原因, 病人中途退出, 或是由于其他原因死亡, 这种情况产生右删失数据. 目前, 已有很多学者关注右删失数据, 可参见文献 [37–40] 等. 近年来, 关于右删失数据非线性小波估计的研究成果也有很多相关文献可以查阅.

在右删失样本中, 令 $\{X_i, 1 \leqslant i \leqslant n\}$ 是独立同分布的生存时间, 具有分布函数 F 和密度函数 f. 令 $\{Y_i, 1 \leqslant i \leqslant n\}$ 是独立同分布的删失时间, 具有分布函数 G. 假设 X_i 和 Y_i 相互独立, 由于随机删失, 我们只能观察到 $Z_i = \min(X_i, Y_i)$ 具有分布函数 H 和 $\delta_i = I(X_i \leqslant Y_i)$. 令 $\tau < \tau_H$ 是固定常数, 其中 $\tau_H = \inf\{x : H(x) = 1\}$, 于是构造小波系数估计量

$$\widehat{\beta}_l = \int \varphi_j(x) I(x \leqslant \tau) \mathrm{d}\widehat{F}_n(x), \qquad \widehat{\beta}_{kl} = \int \psi_j(x) I(x \leqslant \tau) \mathrm{d}\widehat{F}_n(x),$$

其中, $\widehat{F}_n(x)$ 是分布函数 $F(x)$ 的 Kaplan-Meier 估计量, 定义如下

$$\widehat{F}_n(x) = 1 - \prod_{m=1}^{n} \Big(1 - \frac{\delta_{(m)}}{n-m+1}\Big)^{I(Z_{(m)} \leqslant x)},$$

$Z_{(m)}$ 是 Z_i 的次序统计量, $\delta_{(m)}$ 是 $Z_{(m)}$ 对应的右删失因子. 于是, 利用插入法, 可以构造如下形式的未知密度函数 $f(x)$ 的非线性小波估计量

$$\widehat{f}_n(x) = \sum_l \widehat{\beta}_l \varphi_l(x) + \sum_{k=k_0}^{k_1-1} \sum_l \widehat{\beta}_{kl} I(|\widehat{\beta}_{kl}| > \lambda) \psi_{kl}(x), \tag{10}$$

其中 $\lambda > 0$ 和 $k_1 \geqslant 1$ 是光滑参数. Antoniadis 等[41] 构造了右删失数据密度函数非线性小波估计量, 证明了估计量的渐近正态性和估计量 MISE 的最优收敛速度. Li[42,43] 分别给出右删失数据生存时间风险率函数和密度函数非线性小波估计量的渐近正态性和估计量 MISE 的渐近展开式. 薛留根[44] 在完全数据和右删失数据下分别讨论了回归函数小波估计量的强一致收敛速度. Liang 等[45] 证明了右删失相依样本密度函数和风险率函数非线性小波估计量在 Besov 空间中 L_2 误差的一致收敛速度. 潘雄和付宗堂[46] 研究了右删失数据下半参数回归模型参数和函数小波估计量的渐近性质. Liang 和 Qi[47] 讨论了NA条件下右删失样本回归函数非线性小波估计量的渐近正态性. Benatia 和 Yahia[48] 考虑了右删失相依样本回归函数非线性小波估计量的渐近性质. 更多研究可参见文献 [49–53] 等.

§5. 左截断数据非线性小波估计

近年来, 基于艾滋病的传播, 学者们开始关注左截断数据. 在医学研究中, 病人在进行某种疾病的治疗时, 由于某些原因, 病人进行治疗前的相关信息 (体温、血压等) 没有被记录, 这种情况产生左截断数据. 详见文献 [54]. 更多关于左截断数据的研究可参见文献 [55–60] 等.

假设随机变量 $\{X_i, 1 \leqslant i \leqslant n\}$ 具有连续的分布函数 F 和密度函数 f, T_i 是左截断变量, 有连续分布函数 G. 假设 X_i 独立于 T_i. 根据 Lynden-Bell[61] 的思想, 分布函数 F 的乘积限估计量 F_n 定义如下,

$$\widehat{F}_n(x) = 1 - \prod_{X_i \leqslant x} \Big[1 - \frac{1}{nC_n(X_i)}\Big],$$

其中 $C_n(x) = n^{-1} \sum\limits_{i=1}^{n} I(T_i \leqslant x \leqslant X_i)$. 令 $a_H = \inf\{x : H(x) > 0\}$ 和 $b_H = \sup\{x : H(x) < 1\}$. a, b 是满足 $a_G < a < b < b_F$ 两个实数. 构造 $f_1(x) = f(x)I(a < x < b)$ 的局部门限非线性小波估计量

$$\widehat{f}_n(x) = \sum_l \widehat{\rho}_l \varphi_l(x) + \sum_{k=k_0}^{k_1-1} \sum_l \widehat{\rho}_{kl} I(|\widehat{\rho}_{kl}| > \lambda_k), \tag{11}$$

其中 $\lambda_k > 0$ 和 $k_1 \geqslant 1$ 是光滑参数, 小波系数估计量定义如下,

$$\widehat{\rho}_l = \int_0^\infty \varphi_l(x) I(a < x < b) \mathrm{d}\widehat{F}_n(x), \qquad \widehat{\rho}_{kl} = \int_0^\infty \psi_{kl}(x) I(a < x < b) \mathrm{d}\widehat{F}_n(x).$$

Niu 和 Liang[62] 讨论了左截断样本密度函数非线性小波估计量的渐近性质, 并证明了不连续密度函数非线性小波估计量 MISE 的渐近展开式. Niu 和 Xue[63] 研究了左截断相依样本条件密度函数非线性小波估计量的渐近正态性. Cai 和 Liang[64] 讨论了左截断相依样本密度函数非线性小波估计量MISE的渐近展开式. Zou 和 Liang[65] 建立了左截断相依样本密度函数非线性小波估计量在 Besov 空间中全局 L_2 误差的一致收敛速度.

令 $\{X_i, Y_i, T_i, W_i,\, 1\leqslant i\leqslant n\}$ 是随机向量, Y_i是生存时间, 带有分布函数 F, X_i 是与 Y_i 有关的协变量向量, 左截断时间 T_i 带有分布函数 L, 右删失时间 W_i. 在左截断右删失模型里, 可观察的是 $\{X_i, Z_i, T_i, \delta_i\}$, 其中 $Z_i=\min(Y_i, W_i)$ 带有分布函数 H, $\delta_i=I(Y_i\leqslant W_i)$. 根据 Pérez 和 Manteiga[66] 的思想, 分布函数 F 的估计量定义如下,

$$1-\widehat{F}_n(y\,|\,x)=\prod_{i=1}^{n}\left[1-\frac{I(Z_i\leqslant y)\delta_i B_{ni}(x)}{\sum\limits_{j=1}^{n}I(T_j\leqslant Z_i\leqslant Z_j)B_{nj}(x)}\right],$$

其中

$$B_{ni}(x)=K\Big(\frac{x-X_i}{h_n}\Big)\Big/\sum_{j=1}^{n}K\Big(\frac{x-X_j}{h_n}\Big),$$

K 表示核函数, $0<h_n\to 0$ 是带宽序列. 分别定义端点左右支撑 $a_Q=\inf\{y: Q(y)>0\}$ 和 $b_Q=\sup\{y: Q(y)<1\}$. 令固定常数 τ_1 和 τ_2 满足 $a_{L(\cdot|x)}<\tau_1\leqslant\tau_2<b_{H(\cdot|x)}$, 构造如下小波系数估计量

$$\widehat{\omega}_l=\int\varphi_l(y)I(\tau_1\leqslant y\leqslant\tau_2)\mathrm{d}\widehat{F}_n(y\,|\,x),\qquad \widehat{\omega}_{kl}=\int\psi_{kl}(y)I(\tau_1\leqslant y\leqslant\tau_2)\mathrm{d}\widehat{F}_n(y\,|\,x).$$

于是, 利用插入法, 定义如下密度函数非线性小波估计量

$$\widehat{f}_n(x)=\sum_{l}\widehat{\omega}_l\varphi_l(x)+\sum_{k=k_0}^{k_1-1}\sum_{l}\widehat{\omega}_{kl}I(|\widehat{\omega}_{kl}|>\lambda_k). \tag{12}$$

Liang 和 de Uña-Álvarez[67] 研究了左截断右删失相依样本条件密度函数非线性小波估计量 MISE 渐近展开式和估计量的渐近正态性, 并讨论了密度函数不连续性对估计量 MISE 展开式的影响可以忽略.

§6. 缺失数据非线性小波估计

在实际问题中, 很多情况的发生会导致数据缺失. 例如, 在社会调查中, 通过问卷调查或是采访得到的数据, 问卷的遗失或是受访者的不作答都会导致数据缺失. 在临床实验中, 治疗失败、患者搬家或是拒绝继续参与研究都会产生缺失数据. 在工业实验中, 由于各种原因导致部分实验结果记录不完整等. 更多关于缺失数据的研究可参见文献 [68–73] 等.

从缺失机制和方式上可将缺失数据分为三类: 完全随机缺失 (missing completely at random, MCAR)、随机缺失(missing at random, MAR) 和非随机缺失 (missing not at random, MNAR). 在回归分析中, 数据缺失往往分为反应变量缺失和协变量缺失两种情况. 例如, 设 $\{X_i,\, 1\leqslant i\leqslant n\}$ 是协变量, Y_i 是受 X_i 影响的反应变量, 并带有分布函数 F 和密度函数 f. 在实际研究中, 可以得到独立同分布的观察量

$$\{X_i, Y_i, \delta_i\},\qquad i=1,2,\cdots,n,$$

其中, 所有的协变量 X_i 都可被观察, 当 Y_i 缺失时, $\delta_i=0$; 当 Y_i 可被观察时, $\delta_i=1$. 这种情况为反应变量缺失. 类似的可以定义协变量缺失. 在实际应用中出现的数据很多是 MAR 的, 即在给定协变量 X_i 的条件下, Y_i 与 δ_i 相互独立, 即

$$\mathsf{P}(\delta_i=1\,|\,Y_i,X_i)=\mathsf{P}(\delta_i=1\,|\,X_i),$$

上式表明 Y_i 是否缺失与 Y_i 的取值无关, 仅与相应的协变量有关. 详见文献 [74, 75].

Wang 和 Qin [76] 构造了反应变量 Y_i 的分布函数 F 的可逆加权估计量

$$\widehat{F}_{n,W}(y)=\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{\delta_i}{\Delta_n(X_i)}I(Y_i\leqslant y)+\Big[1-\frac{\delta_i}{\Delta_n(X_i)}\Big]F_n(y\,|\,X_i)\Big\},\tag{13}$$

其中 $F_n(y\,|\,x)$ 是条件分布函数 $F(y\,|\,x)=\mathsf{P}(Y\leqslant y\,|\,X=x)$ 的估计量, 定义如下

$$F_n(y\,|\,x)=\frac{\sum\limits_{j=1}^{n}\delta_jI(Y_j\leqslant y)K_1\Big(\dfrac{x-X_j}{h_{1n}}\Big)}{\sum\limits_{j=1}^{n}\delta_jK_1\Big(\dfrac{x-X_j}{h_{1n}}\Big)},\tag{14}$$

其中 K_1 是核函数, $0<h_{1n}\to 0$ 是带宽序列. $\Delta_n(x)$ 是 $\Delta(x)=\mathsf{P}(\delta=1\,|\,X=x)$ 的估计量, 定义如下

$$\Delta_n(x)=\frac{\sum\limits_{j=1}^{n}\delta_jK_2\Big(\dfrac{x-X_j}{h_{2n}}\Big)}{\sum\limits_{j=1}^{n}K_2\Big(\dfrac{x-X_j}{h_{2n}}\Big)},$$

其中 K_2 是核函数, $0<h_{2n}\to 0$ 是带宽序列. 于是可以定义如下形式的小波系数估计量

$$\widehat{\zeta}_l=\int_0^\infty\varphi_l(x)\mathrm{d}\widehat{F}_{n,W}(x),\qquad \widehat{\zeta}_{kl}=\int_0^\infty\psi_{kl}(x)\mathrm{d}\widehat{F}_{n,w}(x).$$

Zou 等 [77] 构造了如下形式的密度函数非线性估计量

$$\widehat{f}_n(x)=\sum_l\widehat{\zeta}_l\varphi_l(x)+\sum_{k=k_0}^{k_1-1}\sum_l\widehat{\zeta}_{kl}I(|\widehat{\zeta}_{kl}|>\lambda_k).\tag{15}$$

Zou 等 [77] 证明了缺失样本密度函数非线性小波估计量 MISE 的渐近展开式和估计量的渐近正态性, 并讨论了密度函数存在有限个间断点时, 估计量 MISE 的渐近展开式仍然成立. Zou 和 Liang [78] 讨论了协变量存在时反应变量随机缺失情况下, 密度函数非线性小波估计量在包含很多不光滑函数的 Besov 空间里全局 L_2 误差的一致收敛速度, 并通过数值模拟计算出估计量平均均方误差 (average mean square error, AMSE) 的最小值, 和对应的最优带宽, 并分析了缺失率和样本容量对估计量的影响. 研究结果表明, 非线性小波估计量的 AMSE 值随着缺失率的增加而增大, 随着样本容量的增加而减少.

§7. 删失模型里删失因子随机缺失数据非线性小波估计

在右删失样本里, 删失因子往往都是可被观察的. 但是, 在实际领域中, 由于各种原因, 导致删失因子随机缺失. 例如, 在生物实验中, 有些科目可能不进行尸检, 以节省开支. 在流行病学研究中, 相关的死亡证明信息可能由于移民而失踪. 在临床实验中, 人们可以区别由感兴趣的疾病引起的死亡和其他原因导致的死亡. 更多信息可参见文献 [79–83] 等.

令随机变量 T_i 表示生存时间, 具有密度函数 f, 随机变量 C_i 表示右删失时间, 假设 T_i 独立于 C_i, 并且 T_i 和 C_i 非负, 我们观察到 $X_i=\min(T_i,C_i)$ 和 $\delta_i=I(T_i\leqslant C_i)$. 令 ξ_i 是随机缺失因子, 如果 δ_i 可被观察, $\xi_i=1$; 如果 δ_i 随机缺失, $\xi_i=0$. 当删失因子 δ_i 随机缺失时, 可观察的变量是

$$\{X_i,\delta_i,\,\xi_i=1\},\quad \{X_i,\delta_i,\,\xi_i=0\},\qquad i=1,2,\cdots,n.$$

在实际应用中, 由于删失因子随机缺失, 传统的估计方法不再适用. Zou 和 Liang[84] 分别利用回归校正法、插补法和可逆概率加权方法, 构造未知密度函数的三种不同的非线性小波估计量. 为了方便构造估计量, 定义 $m(x)=\mathsf{P}(\delta=1\,|\,X=x)$ 和 $\pi(x)=\mathsf{P}(\xi=1\,|\,X=x)$. 在实际问题中, $m(\cdot)$ 和 $\pi(\cdot)$ 都是未知函数, 下面分别给出它们的估计量

$$\widehat{m}_n(x)=\frac{\sum\limits_{j=1}^{n}\xi_j\delta_jK_3\Big(\dfrac{x-X_j}{h_{3n}}\Big)}{\sum\limits_{j=1}^{n}\xi_jK_3\Big(\dfrac{x-X_j}{h_{3n}}\Big)},\qquad \widehat{\pi}_n(x)=\frac{\sum\limits_{j=1}^{n}\xi_jK_4\Big(\dfrac{x-X_j}{h_{4n}}\Big)}{\sum\limits_{j=1}^{n}K_4\Big(\dfrac{x-X_j}{h_{4n}}\Big)},$$

其中, K_3 和 K_4 表示核函数, $0<h_{3n}\to0$ 和 $0<h_{4n}\to0$ 表示带宽序列. 为了保证所有的估计量有意义, 令 $\tau_H=\inf\{x:H(x)=1\}$, 固定的常数 τ 满足 $\tau\leqslant\tau_H$. 利用回归校正法, 定义小波系数的第一种估计量,

$$\widehat{b}_l=\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\varphi_l(X_i)I(X_i\leqslant\tau),$$
$$\widehat{b}_{kl}=\frac{1}{n}\sum_{i=1}^{n}\frac{\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\psi_{kl}(X_i)I(X_i\leqslant\tau),$$

其中, $\widehat{G}_n$ 表示分布函数 G 的估计量, 定义如下

$$\widehat{G}_n(t)=1-\prod_{i:X_i\leqslant t}\Big(\frac{n-R_i}{n-R_i+1}\Big)^{1-\widehat{m}_n(X_i)},$$

$R_i=\sum\limits_{j=1}^{n}I(X_j\leqslant X_i)$ 表示随机变量 X_i 的秩. 接着, 定义小波系数的第二种估计量

$$\widehat{\alpha}_l=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\delta_i+(1-\xi_i)\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\varphi_l(X_i)I(X_i\leqslant\tau),$$

$$\widehat{\alpha}_{kl}=\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_i\delta_i+(1-\xi_i)\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\psi_{kl}(X_i)I(X_i\leqslant\tau),$$

于是, 可以构造密度函数插补非线性小波估计量

$$\widehat{f}_I(x)=\sum_l\widehat{\alpha}_l\varphi_l(x)+\sum_{k=k_0}^{k_1-1}\sum_l\widehat{\alpha}_{kl}I(|\widehat{\alpha}_{kl}|>\lambda_k). \tag{16}$$

最后, 定义小波系数的第三种估计量

$$\widehat{\beta}_l=\frac{1}{n}\sum_{i=1}^{n}\frac{\dfrac{\xi_i\delta_i}{\widehat{\pi}_n(X_i)}+\left[1-\dfrac{\xi_i}{\widehat{\pi}_n(X_i)}\right]\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\varphi_l(X_i)I(X_i\leqslant\tau),$$

$$\widehat{\beta}_{kl}=\frac{1}{n}\sum_{i=1}^{n}\frac{\dfrac{\xi_i\delta_i}{\widehat{\pi}_n(X_i)}+\left[1-\dfrac{\xi_i}{\widehat{\pi}_n(X_i)}\right]\widehat{m}_n(X_i)}{1-\widehat{G}_n(X_i)}\psi_{kl}(X_i)I(X_i\leqslant\tau),$$

于是, 可以构造密度函数可逆概率加权非线性小波估计量

$$\widehat{f}_W(x)=\sum_l\widehat{\beta}_l\varphi_l(x)+\sum_{k=k_0}^{k_1-1}\sum_l\widehat{\beta}_{kl}I(|\widehat{\beta}_{kl}|>\lambda_k). \tag{17}$$

Zou 和 Liang[84] 证明了非线性小波估计量 $\widehat{f}_W$ 的 MISE 渐近展开式和估计量的渐近正态性, 同时讨论了不连续的密度函数的非线性小波估计量 MISE 的渐近展开式仍然成立. 另外, 利用数值模拟分析了估计量平均二次误差 (mean square error, MSE) 随着删失率和缺失率的增加而增大, 随着样本容量的增加而减少. 其中, $\widehat{f}_C$ 的有限样本表现最好, $\widehat{f}_W$ 的有限样本表现最差.

§8.　块门限非线性小波估计

在已有文献中, 大多数非线性小波估计是通过经验小波系数逐项门限构造的. 这些估计量通常可以达到最优收敛速度. Hall 等[28,31] 以组的形式收缩小波系数而不是逐项的形式, 介绍了密度函数和回归函数块门限小波估计量, 并且证明了块门限小波估计量能够达到不带对数项的最优收敛速度. 块门限是介于局部门限与全局门限中间的一种程序. 它能保持或是删除每一级上的特殊封闭的小波系数的块. 基于块门限构造的块门限非线性小波估计量比局部门限非线性小波估计量具有更好的渐近性质, 因为它具有不带额外的对数 ln 项的收敛速度. 但是与全局门限非线性小波估计量相比, 块门限非线性小波估计量具有一个明显的缺点, 就是它依赖于一个不能准确给出的常数 C, C 的选取必须利用经验的方法, 并且依赖被估函数的一致有界性.

Cai[33,85] 研究了回归函数一类块门限小波估计量的渐近全局和局部收敛速度及数值性质. 结果表明, 这些估计量在大范围的 Besov 空间上具有良好的性能. 然而, 所有上述估

计量都是在完全数据下构造的. Li[86] 定义了右删失样本密度函数块门限非线性小波估计量

$$\widehat{f}_n(x)=\sum_j \widehat{\varrho}_j\varphi_j(x)+\sum_{i=0}^{R}\sum_k\sum_{j\in B(k)}\widehat{\varrho}_{ij}\psi_{ij}(x)I(|\widehat{B}_{ik}|>C_0n^{-1}),\tag{18}$$

其中小波系数估计量定义如下

$$\widehat{\varrho}_j=\int\phi_j(x)I(x\leqslant T)\mathrm{d}\widehat{F}_n(x),\qquad \widehat{\varrho}_{ij}=\int\psi_{ij}(x)I(x\leqslant T)\mathrm{d}\widehat{F}_n(x),$$

T 是一个固定的常数, $\widehat{F}_n$ 是随机变量 X 分布函数 F 的 Kaplan-Meier 估计量. R 是光滑参数, $B(k)$ 是块 $\Gamma_{ik}=\{j:(k-1)l\leqslant j\leqslant kl,\ -\infty<k<\infty\}$ 里 j 的集合, l 是无重叠块的长度, C_0 是门限常数, $\widehat{B}_{ik}=l^{-1}\sum\limits_{j\in B(k)}\widehat{\varrho}_{ij}$ 是 $B_{ik}=l^{-1}\sum\limits_{j\in B(k)}\varrho_{ij}$ 的估计量.

Li[86] 讨论了右删失样本密度函数块门限非线性小波估计量在 $B^s_{pq}(M,L)$ 空间里的最优收敛速度. 其中, $B^s_{pq}(M,L)$ 是基于 Besov 空间 B^s_{pq}, 并且满足

$$B^s_{pq}(M,L)=\Big\{f:f\geqslant 0,\int f=1,\ \|f\|_{B^s_{pq}}\leqslant M,\ \mathrm{supp}(f)\subset[-L,L]\Big\},$$

其中 M,L 是有限常数. Shirazi 等[87] 讨论了右删失数据下非参数回归块门限小波估计量的渐近性质. 然而, 其他形式不完全数据下的块门限非线性小波估计还没有文献可以查阅, 这也是未来工作需要努力的方向.

§9. 总结与展望

小波估计方法一直是统计学领域中的研究热点和难点问题, 在其他实际领域有着广泛的应用价值, 本文以小波估计方法在数理统计中的应用为研究对象, 重点介绍小波基本理论、小波估计量的种类、非线性小波估计量门限种类, 以及非线性小波估计方法在完全数据、不完全数据和纵向数据下的理论成果. 由于数据的复杂性、不完全性导致传统的研究方法不再适用, 这就要求学者提出新的思想解决未探究的问题. 小波估计方法的理论研究日益丰富, 将为实际问题的解决提供更多途径. 由于非线性小波估计量比线性小波估计量的收敛速度快, 并且可忽略被估函数不连续性. 在已有的文献中, 更多的是关于非线性小波估计量的研究, 主要内容可以总结为以下五个方面:

(a) 建立被估函数非线性小波估计量 MISE 的渐近展开式;

(b) 证明被估函数非线性小波估计量的渐近正态性;

(c) 讨论被估函数存在有限个不连续点时, 非线性小波估计量 MISE 展开式仍然成立;

(d) 讨论被估函数非线性小波估计量在 Besov 空间里的全局 L_2 误差的一致收敛性;

(e) 证明半参数回归模型中未知参数和连接函数小波估计量的渐近性质, 如相合性、一致收敛性等.

非线性小波估计方法可以估计不光滑的或具有有限个间断点的未知函数, 并且构造出来的非线性小波估计量在包含不连续函数在内的一个大函数空间 B^s_{pq} 空间里可以得到最优收敛速度. 非线性小波估计方法在完全数据的研究有很多文献可以查阅. 然而在实际领域中, 各种原因的发生产生很多复杂数据 (左截断数据、右删失数据、缺失数据、相依数据、纵向数据), 很多学者将非线性小波估计方法应用到复杂数据. 由于数据的不完整性给数据分析和应用带来了很多困难. 根据已有参考文献中关于小波估计方法的研究结果, 可以得到以下三个方面的结论:

(a) 被估函数非线性小波估计量的最优收敛速度比线性小波估计量的最优收敛速度快;

(b) 被估函数的不连续性对非线性小波估计量的影响是可以忽略的, 而经典核估计方法是不成立的;

(c) 被估函数非线性小波估计量的有限样本表现随着不完全数据的删失率、缺失率增大而变差, 随着样本容量增加而变好.

小波估计方法虽然发展迅速, 并且已拥有许多重要的理论结果, 在实际领域的应用也很广泛, 但是还有许多问题尚未解决. 聚焦目前国内外统计学的研究方向, 关于小波理论的未来工作主要集中在以下几个方面:

(a) 在缺失数据下, 关于小波估计方法和理论结果都是在完全独立样本下进行的. 然而, 在实际领域中, 我们经常会遇到很多混合序列, 如 α 混合、ρ 混合、φ 混合, 完全数据下的传统方法不再适用, 这就要求我们发展新的方法;

(b) 随着各领域快速发展, 出现越来越多以曲线和图像形式的无限维函数型数据, 对这类数据的统计推断的关键是降维, 常用的降维方法是对函数型数据进行基展开. 我们可以利用小波基将函数型数据写成基函数的线性组合, 再进行统计推断;

(c) 在实际领域中, 我们会遇到越来越多的高维数据, 因此高维数据与小波理论的结合将会是我们未来的研究方向. 目前这方面的文献还比较少, 可以尝试使用多小波和高维小波对高维数据进行变量选择;

(d) 小波变换的实际应用非常广泛, 尤其是机械工程领域中涉及到非平稳信号分析都可以利用小波变换来处理, 例如滚动轴承故障诊断[88]、旋转机械故障诊断以及地铁车辆转向轴承故障诊断[89] 等. 着眼未来, 小波变换将会为航天设备和高铁动车等重要装备保驾护航, 助力中国制造技术的蓬勃发展.

参 考 文 献

[1] MALLAT S G. A theory for multiresolution signal decomposition: the wavelet representation [J]. *IEEE Trans Pattern Anal Mach Intell*, 1989, **11(7)**: 674–693.

[2] DAUBECHIES I. Ondelettes (Book Reviews: Wavelets and Operators – MEYER Y; Wavelets: Algorithms and Applications – MEYER Y) [J]. *Science*, 1993, **262(5139)**: 1589–1591.

[3] DAUBECHIES I, HEIL C. Ten lectures on wavelets [J]. *Comput Phys*, 1992, **6(6)**: 697.

[4] CHUI C K. *Wavelets: A Tutorial in Theory and Applications* [M]. Boston: Academic Press, 1992.

[5] 陈华丽, 李裕能. 基于小波变换的自适应阈值消噪法 [J]. 电力科学与工程, 2003, **(3)**: 8–10.

[6] 文鸿雁, 张正禄. 非线性小波变换阈值法去噪改进 [J]. 测绘通报, 2006, **(3)**: 18–21.

[7] HÄRDLE W, KERKYACHARIAN G, PICARD D, et al. *Wavelets, Approximation, and Statistical Applications* [M]. New York: Springer, 1998.

[8] DONOHO D L, JOHNSTONE I M. Ideal spatial adaptation by wavelet shrinkage [J]. *Biometrika*, 1994, **81(3)**: 425–455.

[9] DONOHO D L, JOHNSTONE I M. Adapting to unknown smoothness via wavelet shrinkage [J]. *J Amer Statist Assoc*, 1995, **90(432)**: 1200–1224.

[10] DONOHO D L, JOHNSTONE I M, KERKYACHARIAN G, et al. Density estimation by wavelet thresholding [J]. *Ann Statist*, 1996, **24(2)**: 508–539.

[11] KERKYACHARIAN G, PICARD D. Density estimation in Besov spaces [J]. *Statist Probab Lett*, 1992, **13(1)**: 15–24.

[12] WALTER G G. A sampling theorem for wavelet subspaces [J]. *IEEE Trans Inform Theory*, 1992, **38(2)**: 881–884.

[13] DOUKHAN P. *Mixing: Properties and Examples* [M]. New York: Springer-Verlag, 1994.

[14] LIANG H Y. Asymptotic normality of wavelet estimator in heteroscedastic model with α-mixing errors [J]. *J Syst Sci Complex*, 2011, **24(4)**: 725–737.

[15] LEBLANC F. Wavelet linear density estimator for a discrete-time stochastic process: L_p-losses [J]. *Statist Probab Lett*, 1996, **27(1)**: 71–84.

[16] MASRY E. Probability density estimation from dependent observations using wavelets orthonormal bases [J]. *Statist Probab Lett*, 1994, **21(3)**: 181–194.

[17] 李永明, 韦程东. 强混合误差回归函数小波估计的 Berry-Esseen 界 [J]. 数学物理学报, 2009, **29A(5)**: 1453–1463.

[18] 薛留根. 半参数回归模型中小波估计的随机加权逼近速度 [J]. 应用数学学报, 2003, **26(1)**: 11–25.

[19] 胡宏昌, 胡迪鹤. 半参数回归模型小波估计的强相合性 [J]. 数学学报 (中文版), 2006, **49(6)**: 1417–1424.

[20] 刘强, 薛留根. 混合误差下半参数回归模型小波估计的强相合性 [J]. 数学的实践与认识, 2008, **38(10)**: 97–101.

[21] XUE L G, LIU Q. Bootstrap approximation of wavelet estimates in a semiparametric regression model [J]. *Acta Math Sin* (*Engl Ser*), 2010, **26(4)**: 763–778.

[22] LIANG H Y, WANG X Z. Convergence rate of wavelet estimator in semiparametric models with dependent MA(∞) error process [J]. *Chinese J Appl Probab Statist*, 2010, **26(1)**: 35–46.

[23] THALL P F, VAIL S C. Some covariance models for longitudinal count data with overdispersion [J]. *Biometrics*, 1990, **46(3)**: 657–671.

[24] WANG Y G, LIN X, ZHU M. Robust estimating functions and bias correction for longitudinal data analysis [J]. *Biometrics*, 2005, **61(3)**: 684–691.

[25] PANG Z, XUE L G. Estimation for the single-index models with random effects [J]. *Comput Statist Data Anal*, 2012, **56(6)**: 1837–1853.

[26] 刘刚, 刘娟, 刘勇进. 纵向数据变系数回归模型小波估计的渐近性质 [J]. 数学的实践与认识, 2015, **45(8)**: 271–278.

[27] KERKYACHARIAN G, PICARD D, TRIBOULEY K. L_p adaptive density estimation [J]. *Bernoulli*, 1996, **2(3)**: 229–247.

[28] HALL P, KERKYACHARIAN G, PICARD D. On the minimax optimality of block thresholded wavelet estimators [J]. *Statist Sinica*, 1999, **9(1)**: 33–49.

[29] HALL P, PATIL P. Formulae for mean integrated squared error of nonlinear wavelet-based density estimators [J]. *Ann Statist*, 1995, **23(3)**: 905–928.

[30] PATIL P. Nonparametric hazard rate estimation by orthogonal wavelet methods [J]. *J Statist Plann Inference*, 1997, **60(1)**: 153–168.

[31] HALL P, KERKYACHARIAN G, PICARD D. Block threshold rules for curve estimation using kernel and wavelet methods [J]. *Ann Statist*, 1998, **26(3)**: 922–942.

[32] ZHANG S L, ZHENG Z G. Nonlinear wavelet estimation of regression function with random design [J]. *Sci China Ser A*, 1999, **42(8)**: 825–833.

[33] CAI T T. On block thresholding in wavelet regression: adaptivity, block size, and threshold level [J]. *Statist Sinica*, 2002, **12(4)**: 1241–1273.

[34] KERKYACHARIAN G, PICARD D. Density estimation by kernel and wavelets methods: optimality of Besov spaces [J]. *Statist Probab Lett*, 1993, **18(4)**: 327–336.

[35] 薛留根. 混合误差下回归函数小波估计的一致收敛速度 [J]. 数学物理学报, 2002, **22A(4)**: 528–535.

[36] LI L Y, XIAO Y M. Mean integrated squared error of nonlinear wavelet-based estimators with long memory data [J]. *Ann Inst Statist Math*, 2007, **59(2)**: 299–324.

[37] KOUL H, SUSARLA V, VAN RYZIN J. Regression analysis with randomly right-censored data [J]. *Ann Statist*, 1981, **9(6)**: 1276–1288.

[38] STUTE W. The central limit theorem under random censorship [J]. *Ann Statist*, 1995, **23(2)**: 422–439.

[39] DIKTA D. On semiparametric random censorship models [J]. *J Statist Plann Inference*, 1998, **66(2)**: 253–279.

[40] CAI Z W. Estimating a distribution function for censored time series data [J]. *J Multivariate Anal*, 2001, **78(2)**: 299–318.

[41] ANTONIADIS A, GRÉGOIRE G, NASON G. Density and hazard rate estimation for right-censored data by using wavelet methods [J]. *J R Stat Soc Ser B Stat Methodol*, 1999, **61(1)**: 63–84.

[42] LI L Y. Hazard rate estimation for censored data by wavelet methods [J]. *Comm Statist Theory Methods*, 2002, **31(6)**: 943–960.

[43] LI L Y. Non-linear wavelet-based density estimators under random censorship [J]. *J Statist Plann Inference*, 2003, **117(1)**: 35–58.

[44] 薛留根. 完全与删失数据下回归函数小波估计的强一致收敛速度 [J]. 应用数学学报, 2002, **25(3)**: 430–438.

[45] LIANG H Y, MAMMITZSCH V, STEINEBACH J. Nonlinear wavelet density and hazard rate estimation for censored data under dependent observations [J]. *Statist Decisions*, 2005, **23(3)**: 161–180.

[46] 潘雄, 付宗堂. 随机删失半参数回归模型小波估计的渐近性质 [J]. 应用数学学报, 2006, **29(1)**: 68–80.

[47] LIANG H Y, QI Y Y. Asymptotic normality of wavelet estimator of regression function under NA assumptions [J]. *Bull Korean Math Soc*, 2007, **44(2)**: 247–257.

[48] BENATIA F, YAHIA D. Nonlinear wavelet regression function estimator for censored dependent data [J]. *Afr Stat*, 2012, **7(1)**: 391–411.

[49] TRIEBEL H. *Theory of Function Spaces II* [M]. Basel: Birkhäuser, 1992.

[50] VIDAKOVIC B. *Statistical Modeling by Wavelets* [M]. New York: Wiley, 1999.

[51] LI L Y. On the minimax optimality of wavelet estimators with censored data [J]. *J Statist Plann Inference*, 2007, **137(4)**: 1138–1150.

[52] NIU S L. Nonlinear wavelet density estimation with censored dependent data [J]. *Math Methods Appl Sci*, 2012, **35(3)**: 293–306.

[53] LI L Y, LU K W. On rate-optimal nonparametric wavelet regression with long memory moving average errors [J]. *Stat Inference Stoch Process*, 2013, **16(2)**: 127–145.

[54] STRUTHERS C A, FAREWELL V T. A mixture model for time to AIDS data with left truncation and an uncertain origin [J]. *Biometrika*, 1989, **76(4)**: 814–817.

[55] WOODROOFE M. Estimating a distribution function with truncated data [J]. *Ann Statist*, 1985, **13(1)**: 163–177.

[56] KEIDING N, GILL R D. Random truncation models and Markov processes [J]. *Ann Statist*, 1990, **18(2)**: 582–602.

[57] CHAO M T, LO S H. Some representations of the nonparametric maximum likelihood estimators with truncated data [J]. *Ann Statist*, 1988, **16(2)**: 661–668.

[58] STUTE W. Almost sure representations of the product-limit estimator for truncated data [J]. *Ann Statist*, 1993, **21(1)**: 146–156.

[59] LIANG H Y, DE UÑA-ÁLVAREZ J. Conditional quantile estimation with auxiliary information for left-truncated and dependent data [J]. *J Statist Plann Inference*, 2011, **141(11)**: 3475–3488.

[60] 王江峰, 梁汉营, 范国良. 左截断相依数据下非参数回归的局部 M 估计 [J]. 中国科学: 数学, 2012, **42(10)**: 995–1015.

[61] LYNDEN-BELL D. A method of allowing for known observational selection in small samples applied to 3CR quasars [J]. *Monthly Notices Roy Astronom Soc*, 1971, **155(1)**: 95–118.

[62] NIU S L, LIANG H Y. Nonlinear wavelet estimation of conditional density under left-truncated and α-mixing assumptions [J]. *Int J Wavelets Multiresolut Inf Process*, 2011, **9(6)**: 989–1023.

[63] NIU S L, XUE Z A. Asymptotic normality in conditional wavelet density with left-truncated α-mixing observations [J]. *Math Methods Appl Sci*, 2011, **34(5)**: 563–577.

[64] CAI J J, LIANG H Y. Nonlinear wavelet density estimation for truncated and dependent observations [J]. *Int J Wavelets Multiresolut Inf Process*, 2011, **9(4)**: 587–609.

[65] ZOU Y Y, LIANG H Y. Global L_2 error of wavelet density estimator with truncated and strong mixing observations [J]. *Int J Wavelets Multiresolut Inf Process*, 2014, **12(3)**: 1450033 (12 pages).

[66] PÉREZ C I, MANTEIGA W G. Strong representation of a generalized product-limit estimator for truncated and censored data with some applications [J]. *J Nonparametr Stat*, 1999, **10(3)**: 213–244.

[67] LIANG H Y, DE UÑA-ÁLVAREZ J. Wavelet estimation of conditional density with truncated, censored and dependent data [J]. *J Multivariate Anal*, 2011, **102(3)**: 448–467.

[68] WANG Q H, LINTON O, HÄRDLE W. Semiparametric regression analysis with missing response at random [J]. *J Amer Statist Assoc*, 2004, **99(466)**: 334–345.

[69] MÜLLER U U. Estimating the density of a possibly missing response variable in nonlinear regression [J]. *J Statist Plann Inference*, 2012, **142(5)**: 1198–1214.

[70] NIU C Z, GUO X, XU W L, et al. Empirical likelihood inference in linear regression with nonignorable missing response [J]. *Comput Statist Data Anal*, 2014, **79**: 91–112.

[71] WANG Q H, RAO J N K. Empirical likelihood-based inference under imputation for missing response data [J]. *Ann Statist*, 2002, **30(3)**: 896–924.

[72] WANG Q H, RAO J N K. Empirical likelihood-based inference in linear models with missing data [J]. *Scand J Statist*, 2002, **29(3)**: 563–576.

[73] WANG Q H. Probability density estimation with data missing at random when covariables are present [J]. *J Statist Plann Inference*, 2008, **138(3)**: 568–587.

[74] LITTLE R J A, RUBIN D B. *Statistical Analysis with Missing Data* [M]. 2nd ed. New York: Wiley, 2002.

[75] 薛留根. 现代统计模型. 北京: 科学出版社. 2012.

[76] WANG Q H, QIN Y S. Empirical likelihood confidence bands for distribution functions with missing responses [J]. *J Statist Plann Inference*, 2010, **140(9)**: 2778–2789.

[77] ZOU Y Y, LIANG H Y, ZHANG J J. Nonlinear wavelet density estimation with data missing at random when covariates are present [J]. *Metrika*, 2015, **78(8)**: 967–995.

[78] ZOU Y Y, LIANG H Y. Convergence rate of wavelet density estimator with data missing randomly when covariables are present [J]. *Comm Statist Theory Methods*, 2017, **46(2)**: 1007–1023.

[79] SUBRAMANIAN S. Asymptotically efficient estimation of a survival function in the missing censoring indicator model [J]. *J Nonparametr Stat*, 2004, **16(5)**: 797–817.

[80] SUBRAMANIAN S. Survival analysis for the missing censoring indicator model using kernel density estimation techniques [J]. *Stat Methodol*, 2006, **3(2)**: 125–136.

[81] WANG Q H, LIU W, LIU C L. Probability density estimation for survival data with censoring indicators missing at random [J]. *J Multivariate Anal*, 2009, **100(5)**: 835–850.

[82] LI X Y, WANG Q H. The weighted least square based estimators with censoring indicators missing at random [J]. *J Statist Plann Inference*, 2012, **142(11)**: 2913–2925.

[83] WANG Q H, NG K W. Asymptotically efficient product-limit estimators with censoring indicators missing at random [J]. *Statist Sinica*, 2008, **18(2)**: 749–768.

[84] ZOU Y Y, LIANG H Y. Wavelet estimation of density for censored data with censoring indicator missing at random [J]. *Statistics*, 2017, **51(6)**: 1214–1237.

[85] CAI T T. Adaptive wavelet estimation: a block thresholding and oracle inequality approach [J]. *Ann Statist*, 1999, **27(3)**: 898–924.

[86] LI L Y. On the block thresholding wavelet estimators with censored data [J]. *J Multivariate Anal*, 2008, **99(8)**: 1518–1543.

[87] SHIRAZI E, DOOSTI H, NIROUMAND H A, et al. Nonparametric regression estimates with censored data based on block thresholding method [J]. *J Statist Plann Inference*, 2013, **143(7)**: 1150–1165.

[88] 伍建军, 骆建彬. 一种改进的小波分析和 Hilbert 包络的轴承故障 [J]. 机械设计与制造, 2019, **(4)**: 59–62.

[89] 刘建强, 赵东明, 赵楠. 一种改进的地铁车辆转向架轴承故障诊断方法 [J]. 铁道学报, 2018, **40(11)**: 55–61.

Development Review of Wavelet Estimation Method

ZOU Yuye[1,2] FAN Guoliang[2]

([1] *Key Laboratory of Advanced Theory and Application in Statistics and Data Science – MOE, School of Statistics, East China Normal University, Shanghai, 200062, China*)

([2] *College of Economics and Management, Shanghai Maritime University, Shanghai, 201306, China*)

Abstract: Wavelet estimation method has always been one hot and difficult problem in Statistics, and has wide application value in data compression, turbulence analysis, image and signal processing and seismic exploration, etc. The research object of this paper is the application of wavelet estimation method in mathematical statistics, focuses on the basic theory of wavelet estimation method, the types of threshold, and research achievements of the wavelet estimation method under complete data, incomplete data and longitudinal data. Due to the complexity and incompleteness of the data, traditional research methods are no longer applicable. It needs to combine with the characteristics of left truncated data, right censored data, missing data and longitudinal data, use the plug-in, calibration regression, imputation and inverse probability weighting methods. The nonlinear wavelet estimators of estimated functions are constructed, study the asymptotic expansion for mean integral square error (MISE) of nonlinear wavelet estimators and prove the asymptotic normality of estimators. The asymptotic expansions of MISE are still true for the estimated function with finite discontinuous points, and verify the uniform convergence rate of nonlinear wavelet estimators in Besov spaces, which contain unsmoothed functions; as well the wavelet method is used to study the consistency and convergence rate of the parametric and nonparametric parts for the semi-parametric regression models. Finally, the potential development direction of wavelet method is briefly discussed.

Keywords: asymptotic properties; incomplete data; linear wavelet estimation; longitudinal data; nonlinear wavelet estimation

2020 Mathematics Subject Classification: 62G07; 62G20