

# 基于双侧截断的部分线性回归模型的参数估计

沈哲子, 来 鹏\*

南京信息工程大学数学与统计学院, 江苏 南京

收稿日期: 2021年11月23日; 录用日期: 2021年12月17日; 发布日期: 2021年12月24日

---

## 摘 要

受实验观测时长或观测条件导致的数据双侧截断在生存分析等领域十分常见, 故而对这类数据在不同适用模型下的讨论具有研究意义。考虑双侧截断数据下的部分线性模型, 针对观测到数据的逆概率加权, 利用核估计对模型线性部分参数给出估计, 并给出估计量的偏差与方差的渐进表达式, 与常见的模型参数估计方法相比估计值偏差与标准差有一定程度的提升且更准确。通过数值模拟和AIDS病毒感染年龄与潜伏期的相关数据, 验证了该方法的实用性。

## 关键词

双侧截断, 部分线性模型, 加权核估计

---

# Parametric Estimation for Partial Linear Models with Double Truncated Data

Zhezi Shen, Peng Lai\*

School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing Jiangsu

Received: Nov. 23<sup>rd</sup>, 2021; accepted: Dec. 17<sup>th</sup>, 2021; published: Dec. 24<sup>th</sup>, 2021

---

## Abstract

Double truncated data due to the limit of observing time and other objective reasons of the experiments are quite common in the field of survival analysis. So the discussion of this type of data under different models is of great significance. Considering the partial linear model under the two-sided truncated data, based on the inverse probability weighting of the observed data, kernel estimation is used to estimate the linear parameter. Asymptotic expressions for the bias and variance of the estimator are given. Compared with the common method, the bias and variance are

---

\*通讯作者。

**improved. The real data analysis of AIDS virus infection age and incubation proves the practicability of the proposed method.**

## Keywords

**Double Truncation, Partial Linear Model, Weighted Kernel Estimation**

Copyright © 2021 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

双侧截断是一种常出现在生存分析和工业检测上的现象, 双侧截断数据指的是数据的两侧存在由于随机出现的截断点导致响应变量无法被观测, 从而导致样本点缺失的数据集。即当且仅当响应变量  $Y$  落在截断区间  $(L, R)$  内时, 该样本才可以被观测记录。由于响应变量缺失导致的对应协变量的缺失并无规律可循, 而忽视这些有一定占比的缺失数据可能导致估计回归模型时参数有偏, 且拟合效果会随着截断概率的增加而下降, 故而如何在数据存在双侧截断的条件下更好地估计回归模型是一个有实际意义的课题。双侧截断数据的研究最早在天体观测和生存分析领域展开, 从参数法与非参数法出发, 得出了一些有效的估计方法。Efron 和 Petrosian (1999) [1] 针对双侧截断数据的生存分布函数提出了非参数最大似然估计(NPMLE), Shen (2010) [2] 进而研究了 NPMLE 的渐近特性, 并引入了截断分布函数的非参数估计量。Moreira 和 de Uña-Álvarez (2010) [3] 证明了 bootstrap 重采样可用于近似 NPMLE 的抽样分布。当假设的参数模型正确时, 参数法的估计效果要显著强于非参数方法。故而近年来对双侧截断条件下的半参数模型的讨论居多, Shen (2013) [4] 针对半参数变换模型利用 EM 算法给出回归参数估计。进一步的, Shen (2017) [5], Mandel *et al.* (2018) [6] 提出逆概率加权估计法拟合 Cox 回归模型。Rennert 和 Xie (2018) [7] 提出了双截断数据的 Cox 回归模型, 并使用加权估计方程方法来估计回归参数。Ying *et al.* (2019) [8] 提出扩展的 Mann-Whitney 型损失函数, 在响应变量双侧截断时估计线性回归模型中的回归参数。目前已有的研究多集中在生存分析领域常用模型在双侧截断数据情况下的扩展, 但对于一般的常见模型的讨论较少。

部分线性模型是最初由 Engle (1986) [9] 提出的半参数模型, 综合了线性回归的可解释性和非参数回归的稳健性, 有着极广的应用范围。对部分线性回归的估计方法大致有两类: 一是回归样条, 即使用样条函数逼近模型中的非参数部分, 再利用最小二乘法同时估计线性部分参数及非参数部分的样条函数; 二是 Backfitting 估计, 在给定线性部分参数初值条件下估计非参数部分, 再通过极小化估计标准误差求得线性部分参数。然而对于双侧截断数据而言, 已有的估计方法都会因为响应变量的截断导致估计结果有偏。由于双侧截断数据常出现在适用于部分线性模型常见的生存分析, 工业观测等领域, 故而引入适用该类数据的部分线性模型回归方法是有实际价值的。

## 2. 基于双侧截断的部分线性回归模型的参数估计与拟合

本文针对响应变量双侧截断的部分线性回归模型提出一种参数估计方法。

在不考虑数据截断情况下假设全部样本为:  $\tilde{Z} = \{\tilde{X}_i, \tilde{T}_i, \tilde{Y}_i, L_i, R_i, i = 1, \dots, N\}$ , 针对部分线性模型:

$$\tilde{Y} = \tilde{X}^T \beta + g(\tilde{T}) + \tilde{\varepsilon}. \quad (1)$$

对(1)式两侧求期望后直接计算可得:

$$\tilde{Y} - E(\tilde{Y} | T) = \beta^T [\tilde{X} - E(\tilde{X} | T)] + \tilde{\varepsilon}. \quad (2)$$

记  $E(\tilde{Y} | T) = m_{\tilde{Y}}(T)$ ,  $E(\tilde{X} | T) = m_{\tilde{X}}(T)$ , 并且分别将其估计值记为  $\hat{m}_{\tilde{Y}}(T)$ ,  $\hat{m}_{\tilde{X}}(T)$ 。

由于响应变量  $Y$  存在双侧截断, 当且仅当满足  $L_i < \tilde{Y}_i < R_i$  时协变量及响应变量可被观测, 记被观测到的样本为  $Z = \{X_i, T_i, Y_i, U_i, V_i, i = 1, \dots, n\}$ , 其中  $U, V$  为观测到的样本的双侧截断点, 满足  $U_i \leq Y_i \leq V_i$ 。截断后的数据满足:

$$Y = \beta^T X + g(T) + \varepsilon. \quad (3)$$

考虑截断后  $E(Y | T)$  和  $E(X | T)$  的非参数估计。记  $E(Y | T) = m_Y(T)$ ,  $E(X | T) = m_X(T)$ 。

记  $F_1(\cdot | t)$  为给定  $T = t$  时  $\tilde{Y}$  的条件分布函数,  $F_2(\cdot | t)$  为给定  $T = t$  时  $\tilde{X}$  的条件分布函数, 则有  $m_{\tilde{Y}}(t) = \int_{-\infty}^{+\infty} s dF_1(s | t)$ ,  $m_{\tilde{X}}(t) = \int_{-\infty}^{+\infty} s dF_2(s | t)$ 。给定  $T = t$  时  $\tilde{Y}$  落在未被截断区间的概率记为  $\alpha(t)$ , 则有  $\alpha(t) = P(L \leq \tilde{Y} \leq R | T = t) = \int_{-\infty}^{+\infty} H(s) dF_1(s | t)$ , 代表无截断情况下响应变量可以被观测到的条件概率, 其中响应变量可被观测到的概率记为  $H(\cdot)$ , 有  $H(y) = P(L \leq y \leq R)$ 。参照 Moreira C.等(2014) [10]推得截断后响应变量的条件分布函数  $F(y | t) = \alpha(t)^{-1} \int_{-\infty}^y H(s) dF(s | t)$ 。再记  $\alpha^*(t) = \alpha(t)^{-1}$ , 则有  $\alpha^*(t) = \int_{-\infty}^{+\infty} H(s)^{-1} dF(s | t) = E(H(Y)^{-1} | T = t)$ 。

记  $m_Y^*(t) = E(YH(Y)^{-1} | T = t)$ , 故而  $m_Y(t)$  可以表示为:  $m_Y(t) = \frac{m_Y^*(t)}{\alpha_Y^*(t)}$ 。由于  $m^*(t)$  与  $\alpha^*(t)$  均为观

测变量的条件均值, 故二者可以通过 NW 核估计给出估计值。

在双侧截断条件下二者的核估计值分别为:

$$\hat{m}_X^*(T) = \frac{\sum_{i=1}^n K\left(\frac{T_i - T}{h}\right) X_i H_n(Y_i^{-1})}{\sum_{i=1}^{\tilde{n}} K\left(\frac{T_i - T}{h}\right)}, \quad \hat{m}_Y^*(T) = \frac{\sum_{i=1}^n K\left(\frac{T_i - T}{h}\right) Y_i H_n(Y_i^{-1})}{\sum_{i=1}^{\tilde{n}} K\left(\frac{T_i - T}{h}\right)} \quad (4)$$

$$\hat{\alpha}_X^*(T) = \frac{\sum_{i=1}^n K\left(\frac{T_i - T}{h}\right) H_n(Y_i^{-1})}{\sum_{i=1}^{\tilde{n}} K\left(\frac{T_i - T}{h}\right)}, \quad \hat{\alpha}_Y^*(T) = \frac{\sum_{i=1}^n K\left(\frac{T_i - T}{h}\right) H_n(Y_i^{-1})}{\sum_{i=1}^{\tilde{n}} K\left(\frac{T_i - T}{h}\right)} \quad (5)$$

其中  $K(\cdot)$  为核函数,  $h$  为窗宽,  $H_n$  为响应变量被观测到的概率  $H$  的非参数估计。

对于响应变量被观测到概率  $H$  的非参数估计  $H_n$ , Shen (2010) [2]提出  $H_n(y) = \int S_n(dL, dR)$ , 其中  $S_n$  为双侧截断点的联合分布函数的 NPMLE, 并给出了  $S_n$  的迭代估计步骤:

**Step 1.** 设初始估计为  $\phi_i^{(0)} = \sum_{m=1}^n \phi_m^{(0)} I(L_i \leq Y_m \leq R_i)$ , 其中  $\phi_m^{(0)} = \frac{1}{n}, 1 \leq m \leq n$ ;

**Step 2.** 重新对  $S$  给出估计: 令  $\psi_j^{(1)} = \left[ \sum_{i=1}^n \frac{1}{\phi_i^{(0)}} \right]^{-1} \frac{1}{\phi_j^{(0)}}, 1 \leq j \leq n$  并记其累加和为

$$\psi_i^{(1)} = \sum_{m=1}^n \psi_m^{(1)} I(L_m \leq Y_i \leq R_m);$$

**Step 3.** 更新对  $F$  的估计值: 令  $\phi_j^{(1)} = \left[ \sum_{i=1}^n \frac{1}{\psi_i^{(1)}} \right]^{-1} \frac{1}{\psi_j^{(1)}}, 1 \leq j \leq n$  并记其累加和为

$$\phi_i^{(1)} = \sum_{m=1}^n \varphi_m^{(1)} I(L_i \leq Y_m \leq R_i);$$

**Step 4.** 规定收敛条件为  $\max_{1 \leq j \leq n} |\varphi_j^{(k-1)} - \varphi_j^{(k)}| \leq 10^{-6}$ , 重复以上两个迭代步骤直至达到收敛条件。

设在第  $k$  次迭代结束能够满足收敛条件, 则  $S_n$  的估计值为:

$$S_n(L, R) = \sum_{j=1}^n \psi_j^{(k)} I(L_j \leq L, R_j \leq R). \tag{6}$$

由此可知  $H_n(y) = \int S_n(dL, dR) = \sum_{j=1}^n \psi_j^{(k)} I(L_j \leq y \leq R_j)$ 。

至此可计算出  $\hat{m}_Y(T)$ ,  $\hat{m}_X(T)$  分别为:

$$\hat{m}_X(T) = \frac{\hat{m}_X^*(T)}{\hat{\alpha}_X^*(T)}, \quad \hat{m}_Y(T) = \frac{\hat{m}_Y^*(T)}{\hat{\alpha}_Y^*(T)} \tag{7}$$

进一步地, 记  $\bar{Y} = Y - \hat{m}_Y(T)$ ,  $\bar{X} = X - \hat{m}_X(T)$ , 则有:

$$\bar{Y} = \beta^T \bar{X} + \varepsilon. \tag{8}$$

由于双侧截断的缺失机制, 直接对回归系数给出估计存在较大偏差, 故通过逆概率加权对其纠偏, 利用最小二乘法对线性参数  $\beta$  作出估计。

记  $Y_i^* = \frac{\bar{Y}_i}{H_n(Y_i)}$ , 可求得参数  $\beta$  的估计量为:

$$\hat{\beta} = \left( \sum_{i=1}^n \bar{X}_i \bar{X}_i^T \right)^{-1} \sum_{i=1}^n \bar{X}_i Y_i^*. \tag{9}$$

得到线性回归参数的估计量  $\hat{\beta}$  之后, 代入(3)式可以得到非线性部分的拟合值。

为简化表述, 将下述  $\hat{m}_Y(t)$  与  $\hat{m}_X(t)$  统称为  $\hat{m}(t)$ , 给出  $\hat{m}(t)$  的偏差和方差的渐近表达式。定义协变量  $\tilde{T}$  的密度函数为  $f(t)$ ,  $\alpha = P(R \leq Y^* \leq L)$ ,  $\alpha > 0$  为无条件下的观测概率, 同时给定  $\mu(K) = \int t^2 K(t) dt$ ,  $R(K) = \int K(t)^2 dt$ 。为简化表述令  $a_n \sim b_n$  表示实数列  $a_n, b_n$  在  $n \rightarrow \infty$  时满足  $\frac{a_n}{b_n} \rightarrow 1$ 。引入:

$$\sigma^2(x) = E \left[ \left( \tilde{Y} - m(\tilde{T}) \right)^2 H(\tilde{Y})^{-1} \mid T = t \right]. \tag{10}$$

给定如下假设:

- (C1).  $m(t)$  的二阶导  $m''$  连续且在  $t$  的领域内有界;
- (C2). 核函数  $K(\cdot)$  为一关于零对称的密度函数且存在其紧支集;
- (C3).  $n \rightarrow \infty$  的同时  $h \rightarrow 0$ , 且  $nh \rightarrow 0$ ;
- (C4). 条件期望  $E \left[ \left( \tilde{Y} - m(\tilde{T}) \right)^d H(\tilde{Y})^{-1} \mid T = t \right], d = 0, 1, 2$  有界。

基于假设(C1)-(C4), 类似 Moreira C.等(2014) [10]证明过程可以给出  $\hat{m}(t)$  的渐进偏差和方差:

$$Bias(\hat{m}(t)) \sim \frac{1}{2} \mu(K) h^2 \left[ m''(t) f(t) + 2m'(t) f'(t) \right] / f(t) \equiv h^2 B(t), \tag{11}$$

$$Var(\hat{m}(t)) \sim \frac{1}{nh} R(K) \alpha \sigma^2(t) / f(t) \equiv \frac{1}{nh} V(t). \tag{12}$$

### 3. 数值模拟

本节模拟研究文中所提估计方法在有限样本下的实际表现, 比较在响应变量双侧截断的条件下本文提出的估计方法与原始的部分线性模型回归方法的回归效果。对模拟给定的两个模型分别取截断概率为 0.2, 0.4, 可被观测到的样本量分别取  $n = 100, 200, 400$ 。Bootstrap 重抽样数  $B = 1000$ 。记忽略截断条件的参数估计量为  $\hat{\beta}_{PLM}$ , 本文提出的非参数估计法得到的估计量为  $\hat{\beta}_{NP}$ 。对于响应变量双侧截断点的选取分别令其服从均匀分布, 正态分布进行模拟。具体参数设置如下:

模型一: 考虑协变量均为一维变量, 通过模型  $\tilde{Y} = 3\tilde{X} + \sin(2\pi\tilde{T}) + \varepsilon$  产生观测数据, 其中  $\tilde{X} \sim U(-1, 1)$ ,  $\tilde{T} \sim U(0, 1)$ ,  $\varepsilon \sim N(0, 0.01)$ , 随机生成双侧截断点:  $L \sim U(-4, -1)$ ,  $R \sim U(1, 4)$ 。

模型二: 考虑多元协变量, 通过模型  $\tilde{Y} = \beta^T \tilde{X} + \cos(2\pi\tilde{T}) + \varepsilon$  产生观测数据, 其中  $\tilde{X} = (\tilde{X}_1, \tilde{X}_2, \tilde{X}_3)$ ,  $\tilde{X}_i \sim U(-1, 1)$ ,  $(i = 1, 2, 3)$ ,  $\tilde{T} \sim U(0, 1)$ ,  $\beta = (1, 2, -1)^T$ ,  $\varepsilon \sim N(0, 0.01)$ , 随机生成双侧截断点:  $L \sim N(-2, 1)$ ,  $R \sim N(2, 1)$ 。

由表 1 与表 2 可知, 在相同样本量以及相同截断概率的条件下, 本文提出的估计方法相较于传统的估计方法有更小的偏差和较小的估计标准误差, 且 95% 置信区间在真实值附近更加精确。由模型一的参数部分估计结果可以看出, 在双侧截断数据条件下, 已有的部分线性模型估计方法给出的线性部分参数估计结果并没有随着样本数的增加而逼近真实值, 可见数据截断对于该方法的准确性造成一定影响。反观本文提出的非参数方法给出的估计值均为样本数越大, 估计越准确。图 1 与图 2 分别展示了两个模型的非参数部分回归曲线与原始数据曲线(实线)的相近程度。横轴上的线段分布表示了变量的截断情况, 线段越稀疏代表该处被截断从而无法观测到的数据越多。可以看出随着截断后样本数量  $n$  的增加以及样本被观测概率的增长, 传统回归方法(虚线)对原始数据拟合效果提升。而本文提出的对非参数部分的加权核回归(点虚线)几乎与原始数据(实线)重合, 效果明显优于传统方法。

Table 1. Parameter estimation of model 1

表 1. 模型一的参数估计

截断概率	估计量	$n$	Bias	$\widehat{SE}$	95%置信区间
0.2	$\hat{\beta}_{PLM}$	100	0.0234	0.0219	(2.93, 3.02)
		200	0.0039	0.0148	(2.97, 3.03)
		400	0.0281	0.0097	(2.95, 2.99)
	$\hat{\beta}_{NP}$	100	0.0114	0.0044	(2.98, 3)
		200	-0.0106	0.0047	(3, 3.02)
		400	0.0043	0.0030	(2.99, 3)
0.4	$\hat{\beta}_{PLM}$	100	0.0151	0.0199	(2.95, 3.02)
		200	0.0258	0.0118	(2.95, 3)
		400	0.0071	0.0083	(2.98, 3.01)
	$\hat{\beta}_{NP}$	100	-0.0184	0.0127	(2.99, 3.03)
		200	0.0125	0.0023	(2.98, 2.99)
		400	-0.0058	0.0014	(3, 3.01)

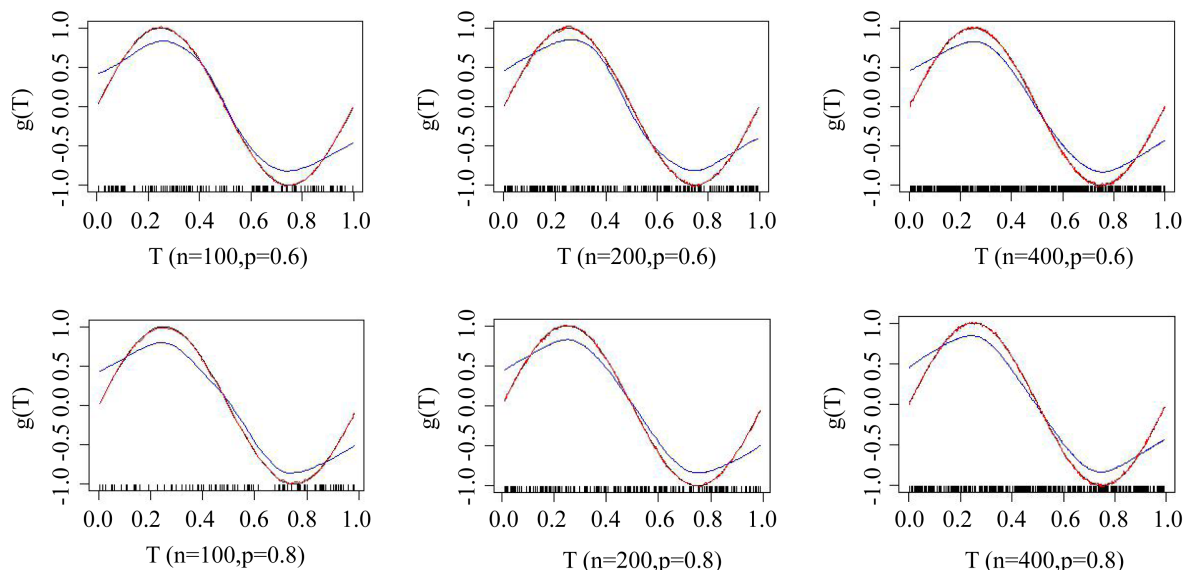


Figure 1. The fitting of two methods to non-parametric part of model 1 under different  $(n, p)$

图 1. 不同  $(n, p)$  取值下两种方法对模型一非参数部分的拟合

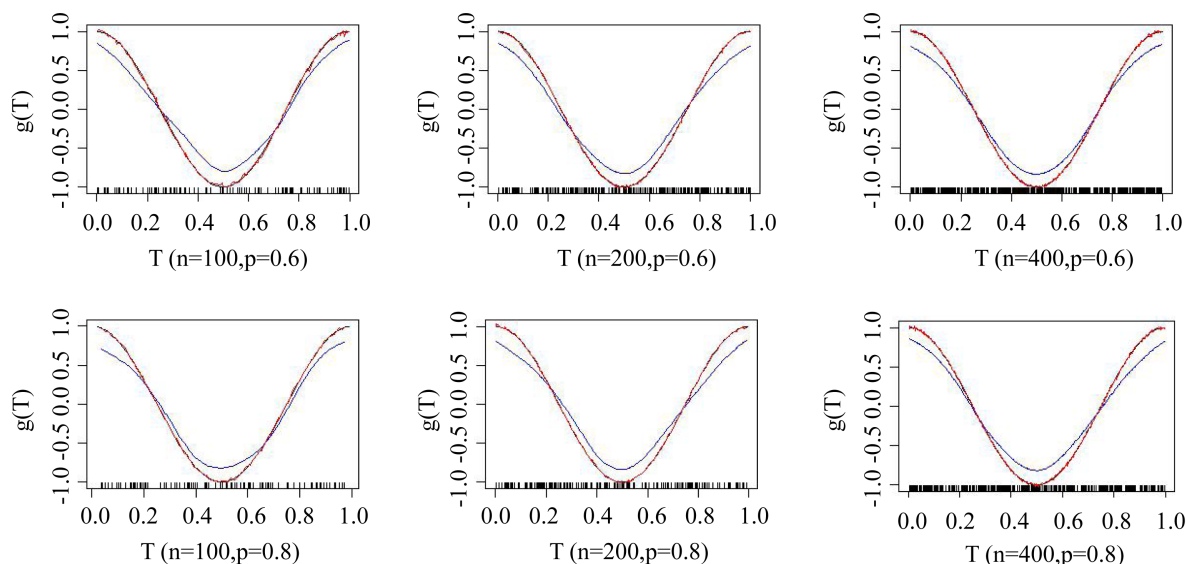
Table 2. Parameter estimation of model 2

表 2. 模型二的参数估计

截断概率	估计量	$n$	$Bias$	$\widehat{SE}$	95%置信区间
0.2	$\hat{\beta}_{PLM}$	100	0.0174	0.0447	(0.96, 1.02), (1.97, 2.03), (-1.01, -0.96)
		200	0.0112	0.0274	(0.98, 1.01), (1.96, 2), (-1.01, -0.97)
		400	-0.0081	0.0140	(0.99, 1.01), (1.97, 1.99), (-1, -0.99)
	$\hat{\beta}_{NP}$	100	-0.0245	0.0204	(0.99, 1.01), (1.99, 2.02), (-1, -0.97)
		200	0.0218	-0.0173	(0.99, 1.02), (1.98, 2.02), (-1.01, -0.98)
		400	-0.0024	0.0149	(1, 1.02), (1.98, 2.01), (-1.01, -0.99)
0.4	$\hat{\beta}_{PLM}$	100	0.0259	0.0396	(0.95, 1.01), (1.95, 2.01), (-1.01, -0.97)
		200	0.0155	0.0183	(0.98, 1.01), (1.97, 1.99), (-1, -0.98)
		400	0.0132	0.0106	(0.98, 0.99), (1.98, 1.99), (-1, -0.99)
	$\hat{\beta}_{NP}$	100	0.0149	0.0119	(0.98, 0.99), (1.97, 1.99), (-0.98, -0.97)
		200	0.0036	0.0173	(0.99, 1.01), (1.98, 2), (-1.01, -0.99)
		400	-0.0027	0.0118	(0.99, 1.01), (1.99, 2.01), (-1.01, -0.99)

#### 4. 实例分析

现将本文提出的方法应用于 Kalbfleisch (1989) 报告的 AIDS 孵化数据共 295 条。数据信息包含患者年龄(AGE), 感染 AIDS 病毒的时长(INF)以及潜伏期(ICUB)。研究目的为判断患者年龄是否会对其病毒潜伏期构成影响。该数据为双侧截断数据, 左侧截断是由于 1982 年之前 HIV 疾病未知, 右侧截断则是实验终止导致。响应变量 ICUB 的截断区间为  $(L, R) = (54 - (INF - 0.5), 138 - (INF - 0.5))$ 。



**Figure 2.** The fitting of two methods to non-parametric part of model 2 under different  $(n, p)$

**图 2.** 不同  $(n, p)$  取值下两种方法对模型二非参数部分的拟合

根据患者年龄将数据划分为三类分别求其年龄与潜伏期的关系, 分别划分为 0~18 为儿童, 19~59 为成人, 60 及以上为老年人, 考虑模型:  $Y = \beta^T X + g(T) + \varepsilon$ 。类似 Liu (2019) 构造变量:  $T = \frac{\text{INF} - E(\text{INF})}{\sqrt{\text{Var}(\text{INF})}}$ 。

利用本文提出的双侧截断数据模型参数估计方法, 给出不同年龄段的病毒潜伏期与年龄的线性回归参数分别为:  $\beta_1 = 3.5$ ,  $\beta_2 = 3.11$ ,  $\beta_3 = 1.49$ 。可得结论: 不同年龄段患者的潜伏期时长受其年龄影响, 且随患者感染年龄的增长该影响逐渐减小。

## 5. 结论

本文基于双侧截断数据, 使用逆概率加权核估计对部分线性模型给出线性部分参数估计, 并给出估计量的偏差与方差的渐进表达式。从数值模拟部分的结果可以看出, 该方法给出的估计值具有更小的偏差与更强的准确性, 且后续对非线性部分的拟合效果有显著的提升, 实例分析进一步证明了该方法的可行性。故而本文提出的针对双侧截断数据下部分线性模型的估计方法在生存分析等领域的适用数据具有实际价值。

## 基金项目

国家自然科学基金面上项目: 超高维复杂数据统计降维研究(11771215), 2018.1~2021.12。

## 参考文献

- [1] Efron, B. and Petrosian, V. (1999) Nonparametric Methods for Doubly Truncated Data. *Journal of the American Statistical Association*, **94**, 824-834. <https://doi.org/10.1080/01621459.1999.10474187>
- [2] Shen, P.S. (2010) Nonparametric Analysis of Doubly Truncated Data. *Annals of the Institute of Statistical Mathematics*, **62**, 835-853. <https://doi.org/10.1007/s10463-008-0192-2>
- [3] Moreira, C. and de Una Alvarez, J. (2010) Bootstrapping the NPMLE for Doubly Truncated Data. *Journal of Nonparametric Statistics*, **22**, 567-583. <https://doi.org/10.1080/10485250903556102>
- [4] Shen, P.S. (2013) Regression Analysis of Interval Censored and Doubly Truncated Data with Linear Transformation

- Models. *Computational Statistics*, **28**, 581-596. <https://doi.org/10.1007/s00180-012-0318-0>
- [5] Shen, P.S. (2017) Semiparametric Analysis of Transformation Models with Dependently Left-Truncated and Right-Censored Data. *Computational Statistics-Simulation and Computation*, **46**, 2474-2487. <https://doi.org/10.1080/03610918.2015.1048879>
- [6] Mandel, M., de Uña-Álvarez, J., Simon, D.K. and Betensky, R.A. (2018) Inverse Probability Weighted Cox Regression for Doubly Truncated Data. *Biometrics*, **74**, 481-487. <https://doi.org/10.1111/biom.12771>
- [7] Rennert, L. and Xie, S.X. (2018) Cox Regression Model with Doubly Truncated Data. *Biometrics*, **74**, 725-733. <https://doi.org/10.1111/biom.12809>
- [8] Ying, Z.L., Yu, W., Zhao, Z.Q. and Zheng, M. (2020) Regression Analysis of Doubly Truncated Data. *Journal of the American Statistical Association*, **115**, 530, 810-821. <https://doi.org/10.1080/01621459.2019.1585252>
- [9] Robert, F., Engle, C.W.J., Granger, J.R. and Weiss, A. (1986) Semiparametric Estimates of the Relation between Weather and Electricity Sales. *Journal of the American Statistical Association*, **81**, 394, 310-320. <https://doi.org/10.1080/01621459.1986.10478274>
- [10] Moreira, C., de Uña, J. and Meira, L.F. (2014) Nonparametric Regression with Doubly Truncated Data. *Computational Statistics and Data Analysis*, **93**, 294-307. <https://doi.org/10.1016/j.csda.2014.03.017>