Summary of Censored Regression: Local Linear Approximations and Their Applications

Presenter: Jaehyeong Ahn

Konkuk University

jayahn0104@gmail.com

January 9, 2021

Table of Contents

Introduction

Introduction

- Fan and Gijbels [1994] suggest a methodology which estimate a regression relationship between response and covariate by using non-parametric regression model, when censored observations (responses) are included in the dataset
- Censoring is a condition in which the value of an observation is only partially known
    - For example, suppose a study is conducted to measure the impact of a heart transplant on survival time
        - If an individual is alive at the end of the study, the true survival time is not observable
        - Then the survival time of this individual is recorded as the last measured value which is called as (right) censored data
- In this paper, only the case of right censoring on response is discussed

Introduction

- When censored observations (responses) are included in the dataset, it is hard to estimate the true relationship between response and covariate
- Fan and Gijbels [1994] propose a methodology for modeling the association, based on censored data
- The methodology implements a proper transformation for censored data and then fit a local linear regression function given the transformed data

Summary of the Methodology

Notation

Model

- $Y = m(X) + \sigma(X)\epsilon$
  - Y : Survival time,
  - X : Associated covariate,
  - $m(\cdot)$: Unknown regression curve,
  - $\sigma(\cdot)$: Conditional variance representing the possible heteroscedacity
- Assume that,
  - $X \perp\!\!\!\perp \epsilon$
  - $E(\epsilon) = 0$, $var(\epsilon) = 1$
- Denote,
  - C: censoring time according to Y
  - $(Y \perp\!\!\!\perp C|X)$
  - $Z = min(Y, C)$, $\delta = I(Y \leq C)$
  - $\{(X_i, Z_i, \delta_i) : i = 1, \cdots, n\}$: a dataset (all observations are ordered according to the X_i's)

Data Example

Simulated Data

X	Y	C	Z	δ
0.0012	0.0867	0.6407	0.0867	1
0.0021	0.2862	0.2844	0.2844	0
0.0026	0.3944	0.1937	0.1937	0
0.0110	0.7629	1.7458	0.7629	1
0.0164	0.9275	0.3252	0.3252	0

Real Data

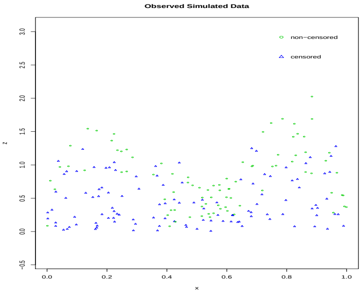
X	Z	δ
12	2.3296	1
13	1.4261	1
14	3.5382	0
15	3.8240	0
18	3.5343	0

Summary of the Methodology

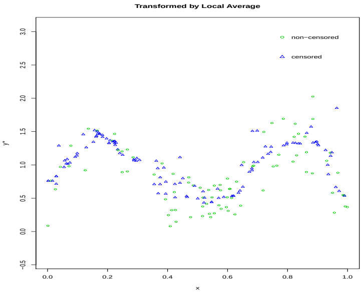
1. **Transformation of the data**
   - For a given integer k and a non-negative weight function K,
   - Replace the observed data $\{(X_i, Z_i, \delta_i)\}$ by $\{(X_i, Y_i^*)\}$, where

$$Y_i^\star = \delta_i Z_i + (1 - \delta_i) \frac{\Sigma_{j:Z_j > Z_i} Z_j K\left(\frac{X_i - X_j}{(X_{i+k} - X_{i-k})/2}\right)\delta_j}{\Sigma_{j:Z_j > Z_i} K\left(\frac{X_i - X_j}{(X_{i+k} - X_{i-k})/2}\right)\delta_j} \tag{1}$$

Summary of the Methodology



(a) Observed data



(b) Transformed by Local Average

Figure: The triangle indicates the censored observations, and the uncensored observations are presented by circle

2 **Application of the local linear regression technique**

- Given the transformed data $\{(X_i, Y_i^*)\}$,
- Fit a local linear regression using the same smoothing parameter k. Namely, compute

$$\hat{m}(x) = \sum_{i=1}^{n} w_i(x) Y_i^\star / \sum_{i=1}^{n} w_i(x), \tag{2}$$

- With

$$w_i(x) \stackrel{\text{def}}{=} K\left(\frac{x - X_i}{\hat{h}_k(x)}\right)[s_{n,2} - (x - X_i) s_{n,1}],$$

- Where

$$s_{n,l} = \sum_{i=1}^{n} K\left(\frac{x - X_i}{\hat{h}_k(x)}\right)(x - X_i)^l, \; l = 0, 1, 2.$$

- $\hat{h}_k(x) = (X_{l+k} - X_{l-k})/2$: an adaptive variable bandwidth (l is the index of the design point X_l closest to x

Summary of the Methodology

3 **Selection of the smoothing parameter k by cross-validation**

- For a given k, Compute below CV function

$$CV(k) = \sum_{i=1}^{n} (Y_i^\star - \hat{m}_{-i}(X_i))^2 \tag{3}$$

- Get $\hat{k}$ which minimize $CV(k)$

$$\hat{k} = \underset{k}{\operatorname{argmin}}\, CV(k), \quad k = 1, ..., [(n-1)/2]$$

  - where $[a]$ is the greatest integer part of a

4 **Calculation of the local linear smoother**

- With $\hat{k}$ as a smoothing parameter, transform the data by (1) and use the smoother (2) to estimate the regression function

Further Explanation of the Methodology

Transformation of the Data

- Let ϕ_1 and ϕ_2 are the transformation functions acting on the uncensored and censored observations
- According to this, we can express the transformation as below

$$
\begin{aligned}
Y^\star &= \phi_1(X, Y) \quad \text{if uncensored} \\
&= \phi_2(X, C) \quad \text{if censored} \\
&= \delta\phi_1(X, Z) \; + \; (1-\delta)\phi_2(X, Z)
\end{aligned}
\tag{4}
$$

- This is regarded as an **"ideal transformation"** because it assumes that ϕ_1 and ϕ_2 are known
- However, those transformation functions typically must be estimated

Transformation of the Data

- To tackle the problem (estimating the transformation functions), Buckley and James [1979] suggested below transformation

$$\phi_1(x, y) = y \text{ and } \phi_2(x, y) = E(Y|Y > y, X = x) \tag{5}$$

  - This transformation can be written as $Y_0^\star = E(Y|\delta, Z, X)$
  - It is closest to the original response in the sense that $E(Y - Y_0^\star)^2 \leq E(Y - Y^\star)^2$; thus it can be regarded as the **"best restoration"**

- But this transformation depends on the unknown regression function leading to an iterative procedure

- And computing the conditional expectation requires strong model assumptions

Transformation of the Data

- Fan and Gijbels [1994] provide the **"estimated transformation"** as follow

$$Y_i^\star = \delta_i Z_i + (1-\delta_i)\frac{\Sigma_{j:Z_j>Z_i} Z_j K\left(\frac{X_i - X_j}{(X_{i+k}-X_{i-k})/2}\right)\delta_j}{\Sigma_{j:Z_j>Z_i} K\left(\frac{X_i - X_j}{(X_{i+k}-X_{i-k})/2}\right)\delta_j} \tag{6}$$

- This transformation estimate the conditional expectation in "best restoration" by using the concept of Nadaraya-Watson (Nadaraya [1964]; Watson [1964]) estimator
- According to this, uncensored observations remain unchanged (i.e., $Y_i^* = Z_i$, if $\delta_i = 1$) and only the censored observations are transformed
- Finally, we get the transformed data $\{(X_i, Y_i^*) : i = 1, \cdots, n\}$

Local Linear Regression Smoothers

- Given the transformed data $\{(X_i, Y_i^*) : i = 1, \cdots, n\}$, estimate the true regression function $m(\cdot)$
- Fix a point x and approximate the unknown function for z in a neighborhood of x as below

$$m(z) \approx m(x) + m^{'}(x)(z - x) \stackrel{\text{def}}{=} a + b(z - x) \tag{7}$$

  - Estimating $m(x)$ is equivalent to estimate the intercept a $(m(x) = E[m(z)|z = x] = a)$
- The local neighborhood and weight are determined by an adaptive variable bandwidth $\hat{h}_k(x)$ and a kernel function K
- the problem of estimating $m(x)$ becomes minimizing w.r.t. a, b

$$\sum_{i=1}^{n} (Y_i^* - a - b(X_i - x))^2 K\left(\frac{x - X_i}{\hat{h}_k(x)}\right)$$

Local Linear Regression Smoothers

- Estimation of $\hat{m}(x)(= \hat{a})$ is provided as follow

$$\hat{m}(x) = \hat{a} = \sum_{i=1}^{n} w_i(x) Y_i^{\star} / \sum_{i=1}^{n} w_i(x), \tag{8}$$

  - With

$$w_i(x) \stackrel{\text{def}}{=} K\left(\frac{x - X_i}{\hat{h}_k(x)}\right)[s_{n,2} - (x - X_i) s_{n,1}],$$

  - Where

$$s_{n,l} = \sum_{i=1}^{n} K\left(\frac{x - X_i}{\hat{h}_k(x)}\right)(x - X_i)^l, \; l = 0, 1, 2.$$

- This estimator possesses several nice properties, including no requirement for boundary modifications, and adaptation to various types of designs such as random and fixed design (Fan [1992])

Asymptotic Result

Adaptive variable bandwidth $\hat{h}_k(x)$

- Let $f_X(\cdot)$ is a marginal probability density function of X
- Following theorem shows that $\hat{h}_k(x)$ behaves approximately as $k/(nf_X(x))$

Theorem 1.

Suppose that $f_X(\cdot)$ is positive and continuous on a compact interval $[a, b]$ and that $k_n \to \infty$ such that $k_n/n \to 0$. Then $\hat{h}_{kn}(x) = [k_n/(nf_X(x)](1 + o_p(1))$ uniformly in $x \in [a, b]$.

- According to Theorem 1, we can observe that $\hat{h}_k(x)$ and $f_X(\cdot)$ are in inverse relationship (when $f_X(\cdot)\downarrow$, $\hat{h}_k(x)\uparrow$)
- Hence this variable bandwidth change the value according to the design density $f_X(\cdot)$

Local linear regression based on the ideal transformation

- Let,
  - $\hat{m}(x; \phi_1, \phi_2)$ be the regression estimator (2) based on the ideal transformation (4),
  - $K(\cdot)$ be a compactly supported PDF with mean 0,
  - $c_K = \int_{-\infty}^{\infty} v^2 K(v)dv, \ d_K = \int_{-\infty}^{\infty} K^2(v)dv$
- Then we have the following result (where, $h_k(x) = k_n/(nf_X(x))$)

Theorem 2.

Suppose that $f_X(\cdot), m^{''}(\cdot)$ and $\sigma^*(\cdot)$ are bounded functions, continuous at the point x, and that $f_X(x) > 0$. If $k_n \to \infty$ such that $k_n/n \to 0$, then, conditionally on the covariates $\{X_1, \cdots, X_n\}$,

$$\sqrt{k_n}(\hat{m}(x; \phi_1, \phi_2) - m(x) - m''(x)c_K h_k^2(x)/2) \to N(0, d_K {\sigma^*}^2(x)),$$

- Theorem 2 shows that local linear regression based on the ideal transformation has asymptotic normality

Local linear regression based on the estimated transformation

- Let,
  - $\hat{m}(x; \hat{\phi}_1, \hat{\phi}_2)$ be the regression estimator (2) based on the estimated transformation (1),
  - Where $\hat{\phi}_1, \hat{\phi}_2$ estimate ϕ_1 and ϕ_2.
- A basic requirement in the consistency result for $\hat{m}(x; \hat{\phi}_1, \hat{\phi}_2)$ is that $\hat{\phi}_1(t, z)$ and $\hat{\phi}_2(t, z)$ are consistent uniformly for t in a neighborhood of x and for z in a chosen interval
- Formally, assume that

$$\beta_n(x) = \max_{j=1,2} \left\{ \sup_{z \in (0, \tau_n),\, t \in (x \pm \tau)} |\hat{\phi}_j(t, z) - \phi_j(t, z)| \right\} = o_p(1) \tag{9}$$

  - with $\tau_n > 0$ and $\tau > 0$

Asymptotic Result

- Let's extend the definition of $\hat{\phi}_j (j = 1, 2)$ as follows to deal with the case of tail $(z > \tau_n)$):

$$\begin{aligned} \hat{\phi}_j(t, z) &= \hat{\phi}_j(t, z), \quad \text{if } z \leq \tau_n \\ &= z \qquad\qquad\quad \text{elsewhere} \end{aligned} \tag{10}$$

- According to this definition, we can express the consistency condition of estimates as follow

$$\kappa_n(x) = \max_{j=1,2} \left\{ \sup_{t \in (x \pm \tau)} E\left(I_{[Z > \tau_n]} |Z - \phi_j(t, Z)| \middle| X = t \right) \right\} = o(1) \tag{11}$$

- Then we have the following result

Theorem 3.

Assume that the conditions of Theorem 1 hold. Then

$$\hat{m}(x;\hat{\phi}_1,\hat{\phi}_2)-\hat{m}(x;\phi_1,\phi_2)=O_p(\beta_n(x)+\kappa_n(x))$$

provided that K is uniformly Lipschitz continuous and has a compact support.

- As a consequence of Theorem 2 and 3, we obtain that $\hat{m}(x;\hat{\phi}_1,\hat{\phi}_2)$ is a consistent estimator of $m(x)$ for any consistent estimators $\hat{\phi}_1$ and $\hat{\phi}_2$

Simulation Result

Simulated Data

Simulation Setting

- We simulated 200 data points from the following model:

$$Y_i = 4.5 - 64X_i^2(1 - X_i)^2 - 16(X_i - 0.5)^2 + 0.25\epsilon_i$$

$$X_i \overset{\text{iid}}{\sim} \text{Uniform}[0, 1], \ \epsilon_i \overset{\text{iid}}{\sim} N(0, 1), X_i \perp\!\!\!\perp \epsilon_i$$

  - Where,
    - $(C_i | X_i = x) \overset{\text{independent}}{\sim} exponential(c(x))$

$$\begin{aligned} c(x) &= 3(1.25 - |4x - 1|), \quad \text{if } \ 0 \leq x \leq .5, \\ &= 3(1.25 - |4x - 3|), \quad \text{if } \ .5 < x \leq 1 \end{aligned}$$

- In this example, approximately 40% of the 200 observations are censored

Simulated Data Example

X	Y	C	Z	δ
0.0012	0.0867	0.6407	0.0867	1
0.0021	0.2862	0.2844	0.2844	0
0.0026	0.3944	0.1937	0.1937	0
0.0110	0.7629	1.7458	0.7629	1
0.0164	0.9275	0.3252	0.3252	0

Simulated Data

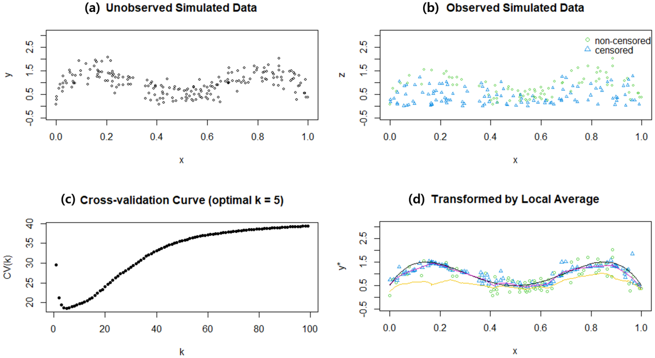


Figure: Simulated Data Set. The triangle indicates the censored observations, and the uncensored observations are presented by circle. The individual panels show (a)unobserved simulated data; (b)observed simulated data; (c) cross-validation curve; (d) data transformed by local average using an optimal $\hat{k} = 5$ and local linear smoothers: black curve-the true regression function, red curve-the smoother based on the transformed data, yellow curve-the smoother based on the observed data in (b)

Real Data

- We used the "Stanford Heart Transplant Data set" which is provided in the "survival" library in R
- This data set include the information of patients who were involved in the clinical (heart transplant) program from 1967 to February, 1989
- In this example, we set 'survival time' as response and 'age' as covariate
- 55 observations out of 175 are censored data

Real Data Example

X	Z	δ
12	2.3296	1
13	1.4261	1
14	3.5382	0
15	3.8240	0
18	3.5343	0

Real Data

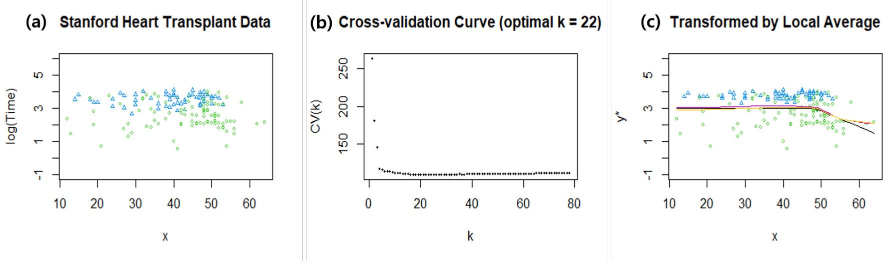


Figure: Stanford Heart Transplant Data Set. The triangle indicates the censored observations, and the uncensored observations are presented by circle. The individual panels show (a)log-survival time plotted against age; (b)cross validation curve; (c) data transformed by local average using an optimal $\hat{k} = 22$ and local linear smoothers: black curve-the suggested relationship, red curve-the smoother based on the transformed data, yellow curve-the smoother based on the observed data in (a)

References

- Jonathan Buckley and Ian James. Linear regression with censored data. Biometrika, 66(3):429–436, 1979.
- William S Cleveland. Robust locally weighted regression and smoothing scatterplots. Journal of the American statistical association, 74(368):829–836, 1979.
- Jianqing Fan. Design-adaptive nonparametric regression. Journal of the Amer- ican statistical Association, 87(420):998–1004, 1992.
- Jianqing Fan and Irene Gijbels. Censored regression: local linear approximations and their applications. Journal of the American Statistical Association, 89 (426):560–570, 1994.
- Elizbar A Nadaraya. On estimating regression. Theory of Probability Its Applications, 9(1):141–142, 1964.

References

- Charles J Stone. Consistent nonparametric regression. The annals of statistics, pages 595–620, 1977.
- Geoffrey S Watson. Smooth regression analysis. Sankhy a: The Indian Journal of Statistics, Series A, pages 359–372, 1964.

Q&A