Nonparametric variable selection via sufficient dimension reduction for cross-sectional survival data without follow-up

Ming-Yueh Huang

Institute of Statistical Science, Academia Sinica

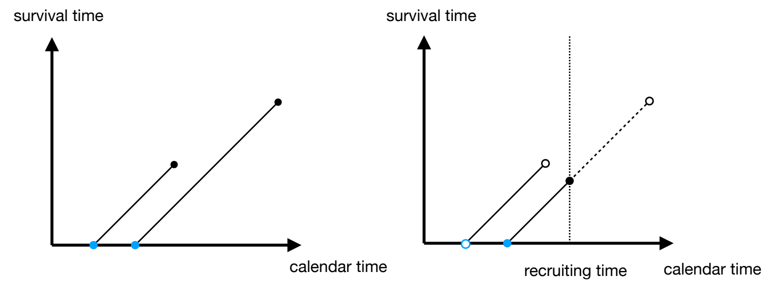
December 12, 2020

Regression on Time-to-Event Data

- T^0: failure time of interest
  - Duration time from an initial event to a failure event
- X^0: baseline covariate at onset
- Parameter of interest: $S_{T^0}(t \mid X^0 = x)$

Sampling Mechanisms

- Incident vs. prevalent sampling



- Cross-sectional data: left truncation
- Without follow-up: fully right censoring

Data Structures

- Prevalent data without follow-up: (A, X)
  - A^0: truncation time
  - $(A, X) \sim (A^0, X^0) \,|\, T^0 > X^0$
- Incident and prevalent covariate data: X and X^0
  - X and X^0 are independent
- Problems
  - Only biased sample of (T^0, X^0) is available.
  - Data with no follow-up: $S_{T^0}(t \,|\, X^0 = x)$ is not identifiable.
- Parametric models are often used to identify the covariate effects.

Length-Biased Data

- A^0 is (improper) uniformly distributed.
- Yamaguchi (2003)
  - $\ln T^0 = -\beta_0^{\mathrm{T}} X^0 + \varepsilon_0 \Rightarrow \ln A = -\beta_0^{\mathrm{T}} X + \varepsilon^*$
- Oakes and Dasu (1990), Chan et al. (2012)
  - Proportional mean residual life model: $\mathrm{E}(T^0 - t \mid T^0 > t, X^0 = x) = m(t) \exp(\beta_0^{\mathrm{T}} x)$
  - $\lambda_A(t \mid x) = \{1/m(t)\} \exp(-\beta_0^{\mathrm{T}} x) = m^*(t) \exp(-\beta_0^{\mathrm{T}} x)$
- Chan (2013)
  - $\ln \mathrm{E}[T^0 \mid X^0] = \alpha_0 + \beta_0^{\mathrm{T}} X \Rightarrow f_X(x) = \frac{\exp(\alpha_0 + \beta_0^{\mathrm{T}} x)}{\mathrm{E}(T^0)} f_{X^0}(x)$
  - $\ln\{f_X(x)/f_{X^0}(x)\} = \alpha^* + \beta_0^{\mathrm{T}} x$

More General Modeling

- $A^0 \perp\!\!\!\perp (T^0, X^0)$
- Chen and Chiang (2018)
  - General truncation distribution is allowed.
  - General single-index model is considered: $S_{T^0}(t \,|\, x) = S_0(t, \beta_0^{\mathrm{T}} x)$
  - For prevalent data without follow-up:
    $f_A(t \,|\, x) = S_0(t, \beta_0^{\mathrm{T}} x) f_{A^0}(t) / \int_0^\infty S_0(u, \beta_0^{\mathrm{T}} x) f_{A^0}(u) du$
    $\Rightarrow f_A(t \,|\, x) = f(t, \beta_0^{\mathrm{T}} x)$
  - For incident and prevalent covariate data
    $f_X(x)/f_{X^0}(x) = \int_0^\infty S_0(u, \beta_0^{\mathrm{T}} x) f_{A^0}(u) du / \mathrm{P}(T^0 > A^0)$
    $\Rightarrow \ln\{f_X(x)/f_{X^0}(x)\} = g(\beta_0^{\mathrm{T}} x)$

Challenges

- Is the single-index model assumed correctly?
    - Model diagnosis is required.
    - Difficulty: $S_0(t, v)$ is not identifiable.
    - More general models?
- How to characterize/screen the covariate effects in more general semiparametric/nonparametric models.

Key Idea

- Connection between (A, X) and (A^0, T^0, X^0)
  - $f_A(t \mid x)$
  - $\ln\{f_X(x)/f_{X^0}(x)\}$
- Important finding:
  - $S_{T^0}(t \mid X^0 = x)$ is only partially dellivered.
  - Central subspace of $S_{T^0}(t \mid X^0 = x)$ can be fully delivered.

Sufficient Dimension Reduction

- The central subspace $\mathcal{S}_{T^0|X^0} = \mathrm{span}(B_0)$ is the smallest linear subspace such that

  $$T^0 \perp\!\!\!\perp X^0 \,|\, B_0^{\mathrm{T}} X^0,$$

  where B_0 is a $p \times d_0$ index coefficient matrix.

- Equivalently, $\mathcal{S}_{T^0|X^0}$ is the smallest linear subspace such that

  $$S_{T^0}(t \,|\, X^0 = x) = S(t, B_0^{\mathrm{T}} x) \text{ for some link function } S.$$

Submodels

- The structural dimension d_0 (number of linear indices) is also to be determined.
- In fact, SDR is a series of nested multiple index models.
  - $d_0 = p$: fully nonparametric regression
  - $d_0 = 1$: single-index model
  - $d_0 = 0$: $T^0 \perp\!\!\!\perp X^0$
- $\mathrm{e}_\ell^{\mathrm{T}} B_0 \equiv 0 \Leftrightarrow X_\ell^0$ has no covariate effect on T^0.
- Parameters of interest: d_0, B_0, and $\mathcal{A}_0 = \{\ell : \|\mathrm{e}_\ell^{\mathrm{T}} B_0\| \neq 0\}$.

Prevalent Data without Follow-Up

- Observed variables: $(A, X) \sim (A^0, X^0) \,|\, T^0 > X^0$
- Under $A^0 \perp\!\!\!\perp (T^0, X^0)$,

$$f_A(t \,|\, x) = \frac{S_{T^0}(t \,|\, X^0 = x) f_{A^0}(t)}{\int_0^\infty S_{T^0}(u \,|\, X^0 = x) f_{A^0}(u) du}.$$

- The central subspace $\mathcal{S}_{A|X} = \mathcal{S}_{T^0|X^0}$.
- Random sample $\{(A_i, X_i)\}_{i=1}^n$
- Existing SDR methods can be directly applied.

Semiparametric Cross-Validation Criterion

- Huang and Chiang (2017)
- The semiparametric cross-validation criterion is defined as

$$\textsc{cv}_A(d, B, h) = \frac{1}{n}\sum_{i=1}^{n}\int\{1(A_i \le t) - \widehat{F}_A^{-i}(t \mid B^{\mathrm{T}}X_i; B)\}^2 d\widehat{F}_A(t),$$

  - $\widehat{F}_A(t \mid v; B) = \frac{\sum_{i=1}^{n} 1(A_i \le t)\mathcal{K}_h(B^{\mathrm{T}}X_i - v)}{\sum_{i=1}^{n} \mathcal{K}_h(B^{\mathrm{T}}X_i - v)}$,
  - The supersript $-i$ denotes the estimator based on data with the ith subject being deleted,
  - $\widehat{F}_A(t) = n^{-1}\sum_{i=1}^{n} 1(A_i \le t)$.
- $(\widehat{d}, \widehat{B}, \widehat{h}) = \mathrm{argmin}\textsc{cv}_A(d, B, h)$.

Selection of Baseline Significant Covariates

- The parametrization $B = (I_d, C^{\mathrm{T}})^{\mathrm{T}}$ requires d baseline significant covariates for each working dimension d.
- Minimizing $\mathrm{CV}_A(d, B, h)$ w.r.t. all the permutations of $\{X_1, \ldots, X_p\}$.
  - At least $\binom{p}{d}$ minimization problems to be solved
- Starting from an initial estimator $\check{B}$
  - Calculate the projection matrix $\check{P} = \check{B}(\check{B}^{\mathrm{T}}\check{B})^{-1}\check{B}^{\mathrm{T}}$.
  - Calculate the L^2-norms of column vectors of $\check{P}$.
  - Choose the covariates corresponding to the column vectors with first d large norms.

Penalization

- Screening of zero row vectors of $\widehat{B}$
- Consistent selection (oracle property)
- We combine the group LASSO of Yuan and Lin (2006) and the adaptive LASSO of Zou (2006).

$$\mathrm{CV}_{A,\lambda}(B) = \mathrm{CV}_A(\widehat{d}, B, \widehat{h}) + \lambda \sum_{\ell=1}^{p-\widehat{d}} \frac{\|\mathrm{e}_\ell^{\mathrm{T}} B\|}{\|\mathrm{e}_\ell^{\mathrm{T}} \widehat{B}\|},$$

where λ is a tuning parameter.

- $\widehat{B}_\lambda = \mathrm{argmin} \mathrm{CV}_{A,\lambda}(B)$.

Tuning Parameter

- We modify the generalized information criterion of Zhang et al. (2010) to the following BIC-type criterion:

$$\mathrm{CV}_A(\widehat{d}, \widehat{B}_\lambda, \widehat{h}) + \frac{\log n}{n} |\widehat{\mathcal{A}}_\lambda| \mathrm{CV}_A(\widehat{d}, \widehat{B}, \widehat{h}),$$

where $\widehat{\mathcal{A}}_\lambda = \{\ell : \|\mathrm{e}_\ell^{\mathrm{T}} \widehat{B}_\lambda\| \neq 0\}$, and denote the minimizer as $\widehat{\lambda}$.

Incident and Prevalent Covariate Data

- Observed variables: X and X^0 with $X \perp\!\!\!\perp X^0$
- Under $A^0 \perp\!\!\!\perp (T^0, X^0)$,

$$f_X(x) = \frac{\int_0^\infty S_{T^0}(u \mid X^0 = x) f_{A^0}(u) du}{\mathrm{P}(T^0 > A^0)} f_{X^0}(x).$$

- Let

$$m(x) \triangleq \ln\{f_X(x)/f_{X^0}(x)\} = \ln \frac{\int_0^\infty S_{T^0}(u \mid X^0 = x) f_{A^0}(u) du}{\mathrm{P}(T^0 > A^0)}.$$

Case-Control Study

- Designed variables:

$$(D, Z) = \begin{cases} (0, X^0) & \text{if subject belongs to the incident cohort,} \\ (1, X) & \text{if subject belongs to the prevalent cohort.} \end{cases}$$

- $f_Z(z \mid D = 1) = f_X(z)$ and $f_Z(z \mid D = 0) = f_{X^0}(z)$.
- $\begin{cases} \mathrm{P}(D = 1 \mid Z = z) = \frac{\exp\{m(z)\}\mathrm{P}(D=1)}{\exp\{m(z)\}\mathrm{P}(D=1)+\mathrm{P}(D=0)}, \\ m(z) = \ln \frac{\mathrm{P}(D=1 \mid Z=z)\mathrm{P}(D=0)}{\mathrm{P}(D=0 \mid Z=z)\mathrm{P}(D=1)}. \end{cases}$
- $\mathcal{S}_{D|Z} = \mathrm{span}(B_0)$ is the smallest subspace s.t. $m(x) = g(B_0^{\mathrm{T}} x)$.
- Under some mild conditions, $\mathcal{S}_{D|Z} = \mathcal{S}_{T^0|X^0}$.

Estimation

- $\text{CV}_D(d, B, h) = \frac{1}{n}\sum_{i=1}^{n_0+n_1}\{D_i - \widetilde{\pi}^{-i}(B^{\mathrm{T}}Z_i; B)\}^2$.
  - $\widetilde{\pi}(v; B) = \frac{\sum_{i=1}^{n} D_i \mathcal{K}_h(B^{\mathrm{T}}X_i - v)}{\sum_{i=1}^{n} \mathcal{K}_h(B^{\mathrm{T}}X_i - v)}$.
- $(\widetilde{d}, \widetilde{B}, \widetilde{h}) = \text{argmin}\text{CV}_D(d, B, h)$.
- $\text{CV}_{D,\lambda}(B) = \text{CV}_D(\widetilde{d}, B, \widetilde{h}) + \lambda \sum_{\ell=1}^{p-\widehat{d}} \frac{\|\mathrm{e}_\ell^{\mathrm{T}} B\|}{\|\mathrm{e}_\ell^{\mathrm{T}} \widetilde{B}\|}$.
- $\widetilde{B}_\lambda = \arg\min_B \ \text{CV}_{D,\lambda}(B)$.
- $\widetilde{\lambda} = \text{argmin} \ \text{CV}_D(\widetilde{d}, \widetilde{B}_\lambda, \widetilde{h}) + \frac{\log n}{n}|\widetilde{\mathcal{A}}_\lambda|\text{CV}_D(\widetilde{d}, \widetilde{B}, \widetilde{h})$.

Simulations

- $X = (X_1, \ldots, X_{10})^{\mathrm{T}} \sim N(0, I_{10})$
- $\beta_{01} = (1, 0, 0, 0, 0, 0, 0, 1, 1, 1)^{\mathrm{T}}$, $\beta_{02} = (1, 0, 0, 0, 0, 0, 0, 1, 1, 0)^{\mathrm{T}}$, and $\beta_{03} = (0, 1, 0, 0, 0, 0, 0, 0, 1, -1)^{\mathrm{T}}$
- $\varepsilon \sim N(0, 0.05^2)$

M1. $T^0 = [1/\{1 + \exp(1 - \beta_{01}^{\mathrm{T}} X^0)\}] \exp(\varepsilon)$

M2. $T^0 = (1/[1 + \exp\{1 - (\beta_{02}^{\mathrm{T}} X^0)(\beta_{03}^{\mathrm{T}} X^0)\}]) \exp(\varepsilon)$

- $A^0 \sim \mathrm{Unif}(0, c_{\mathrm{U}})$ or $\mathrm{Beta}(2, c_{\mathrm{B}})$
- $P(T^0 \geq A^0)$: 0.2, 0.4, 0.6

Simulations

- When the sample size increases,
  - the proportions of $\widehat{d} = d_0$ tend to one,
  - the proportions of selecting significant covariates tend to one,
  - the proportions of insignificant covariates tend to zero,
  - the accuracy measures of $\widehat{B}$ and $\widehat{B}_\lambda$ tends to zero.
- The penalized estimator can not guarantee smaller accuracy measure but mostly lead to smaller standard error of the accuracy measure.
- Same conclusions for $(\widetilde{d}, \widetilde{B})$ and $\widetilde{B}_{\widetilde{\lambda}}$
- The same increasing size on n_1 usually leads to better performance than that on n_0.

National Comorbidity Survey Replication Data

- The survey was conducted in 2001-2002.
  - 1010 English-speaking household residents aged 18+ years old
- Childhood adversities $\rightarrow$ durations of adult mental disorders
- T^0: duration between suicidal thoughts
- A^0: time from the last event to recruitment
- Baseline covariates X^0:
  - age of last suicidal thoughts (age),
  - family structure (fs),
  - gender ($gender$),
  - status of ever using marijuana or hashish ($drug$)

National Comorbidity Survey Replication Data

- Chen and Chiang (2018)
  - PLISE: $age - 0.58fs + 0.07gender + 1.39drug$
  - Rank correlation estimator: $age - 1.33fs + 0.15gender + 1.33drug$
- $\widehat{B}_{\widehat{\lambda}}^{\mathrm{T}} X = age + 0.083drug$
  - The single-index model is adequate.
  - fs and $gender$ have no covariate effects.

Worcester Heart Attack Study Data

- Approximately 23% random sample from the cohort years 1997, 1999, and 2001 (Hosmer et al. 2008)
- T^0: survival following admission to a hospital after AMI
- Baseline covariates X^0:
  - age (age)
  - body mass index (BMI) at hospital admission
  - gender ($gender$)
- Patients
  - prevalent: admitted to hospitals before April 1, 1999 and were still followed at this date, $n_1 = 151$
  - incident: admitted to hospitals after April 1, 1999, $n_0 = 300$

Worcester Heart Attack Study Data

- Chen and Chiang (2018)
  - Rank correlation estimator:
    $age - 0.20 gender - 0.03 BMI + 0.30 BMI^2$
- $\widetilde{B}_{\widetilde{\lambda}}^{\mathrm{T}} X = age$
  - The single-index model is adequate.
  - age is the only significant covariate.

Conclusion

- For cross-sectional data without follow-up, we showed that the central subspace can be fully delivered.
- Instead of assuming particular models, we select the correct model in a series of nested models which contains the fully nonparametric regression.
- The central subspace can help detect redundant covariates.

Future Work

- Combination of two types of data
  - Optimal weights for combined criteria
- High-dimensional covariates
  - Pre-screening