A Simple Method for Estimating Interactions Between a Treatment and a Large Number of Covariates

Purposal

- Model interactions between treatment and covariates
- Focus on a subset of patients rather than overall study population
- “modified covariates” idea can be used to hypothesize test for determining which of a set of covariates interact with a treatment variable.

Common used approaches

- 1. compare a subset of patients' treatment and control arms in different subgroups defined a priori

→focus on interaction between treatment and a dichotomized covariate

→Defects: false positive findings and simple interactions

- 2. use the regression model with the product of the binary treatment indicator and covariates

→Defects: hard to detect the interaction

→reduce the number of covariates interacting with treatment

Definition

- $T = \pm 1$: the binary treatment indicator
- $Y^{(1)}$ and $Y^{(-1)}$ the potential outcome if the patient received treatment $T = \pm 1$
- Observed data $\{(Y_i, T_i, \boldsymbol{Z}_i)\}$: *N* i.i.d. copies of $\{(Y, T, \boldsymbol{Z})\}$
- ***W***(·): $R^q \rightarrow R^p$: a p-dimensional functions of baseline covariates ***Z,*** and denote $\boldsymbol{W}(\boldsymbol{Z}_i)$ by $\boldsymbol{W}_{\mathrm{i}}$

Assumption

- The treatment is randomly assigned to a patient, i.e. T and $\boldsymbol{Z}$ are independent.
- For simplicity, $P(T = 1) = P(T = -1) = 1/2$

Continuous Response Model

- Y is a continuous response and linear regression model is

Model 1 $$Y = \beta_0' \mathbf{W}(\mathbf{Z}) + \boldsymbol{\gamma}_0' \mathbf{W}(\mathbf{Z}) \cdot T/2 + \epsilon,$$ W(Z) includes intercept

- where ϵ is the mean zero random error.
- $\gamma_0' W(Z) \cdot T$ models treatment effect across the population
- $\gamma_0' W(Z)$ can be used to indentify the subset of patients who may or may not benefit from the treatment

Estimator γ_0

Method 1: $$\Delta(\mathbf{z}) = \mathrm{E}(Y^{(1)} - Y^{(-1)} | \mathbf{Z} = \mathbf{z}) = \boldsymbol{\gamma}_0' \mathbf{W}(\mathbf{z}),$$

- $\gamma_0' W(Z)$ measures the causal treatment effect
- γ_0 can be estimated by OLS method

- *Method 2:* $$\mathrm{E}(2YT | \mathbf{Z} = \mathbf{z}) = \Delta(\mathbf{z}),$$
- γ_0 can be estimated by minimizing

$$\frac{1}{N} \sum_{i=1}^{N} (2Y_i T_i - \boldsymbol{\gamma}' \mathbf{W}_i)^2.$$

Estimator γ_0

- 1. estimators are consistent for γ_0
- 2. Even when the model is misspecified, the modified outcome estimators still converge to the same nonrandom limit γ^*
- 3. score $W(z)'\gamma^*$ is a sensible estimator for the interaction effect to find z by minimizing

$\gamma' W(z)$

$$E\{\Delta(\mathbf{Z}) - f(\mathbf{Z})\}^2 \text{ subject to } f \in \mathcal{F} = \{\boldsymbol{\gamma}'\mathbf{W}(\mathbf{z}) | \boldsymbol{\gamma} \in R^p\},$$

Modified Covariate Method

- Consider model

$$Y = \gamma_0' \mathbf{W}(\mathbf{Z}) \cdot T/2 + \epsilon,$$

- Estimator $\hat{\gamma}$ by minimizing

$T_i^2 = 1$

$$\frac{1}{N} \sum_{i=1}^{N} \left\{ Y_i - \gamma' \frac{\mathbf{W}_i \cdot T_i}{2} \right\}^2 = \frac{1}{4N} \sum_{i=1}^{N} \{2Y_i T_i - \gamma' \mathbf{W}_i\}^2 ,$$

T_i^2

Proposal

- 1. Modify the covariate

$$Z_i \rightarrow \mathbf{W}_i = \mathbf{W}(\mathbf{Z}_i) \rightarrow \mathbf{W}_i^* = \mathbf{W}_i \cdot T_i/2$$

- 2. Perform appropriate regessions without intercept

$$Y \sim \boldsymbol{\gamma}_0' \mathbf{W}^* \qquad (W_i^*, Y_i) \qquad (5)$$

- 3. $\hat{\gamma}'\mathrm{W}(\mathrm{z})$ can be used to stratify patients for their treatment

Comparison of Binary and Survival Response

	Binary Response	Survival Response
Y	binary response	$(X,\delta)=\{\min(\tilde{X},C), 1(\tilde{X} < C)\}$
fitting model	logisitic model $$P(Y=1\mid \mathbf{Z},T)=\frac{\exp(\boldsymbol{\gamma}_0'\mathbf{W}^*)}{1+\exp(\boldsymbol{\gamma}_0'\mathbf{W}^*)}.$$	Cox model $$\lambda(t\mid \mathbf{Z},T)=\lambda_0(t)e^{\boldsymbol{\gamma}'\mathbf{W}^*},$$
the expression of $\Delta(\boldsymbol{z})$ in specified model	$$\begin{aligned}\Delta(\mathbf{z})&=P(Y^{(1)}=1\mid \mathbf{Z}=\mathbf{z})-P(Y^{(-1)}=1\mid \mathbf{Z}=\mathbf{z})\\&=\frac{\exp\{\boldsymbol{\gamma}_0'\mathbf{W}(\mathbf{z})/2\}-1}{\exp\{\boldsymbol{\gamma}_0'\mathbf{W}(\mathbf{z})/2\}+1}\end{aligned}$$	$$\Delta(\mathbf{z})=\frac{E\{\Lambda_0(\tilde{X}^{(1)})\mid \mathbf{Z}=\mathbf{z}\}}{E\{\Lambda_0(\tilde{X}^{(-1)})\mid \mathbf{Z}=\mathbf{z}\}}=\exp\{-\boldsymbol{\gamma}_0'\mathbf{W}(\mathbf{z})\}$$ $$\Lambda_0=\int_0^t \lambda_0(u)\,du$$

Comparison of Binary and Survival Response

<table>
<tr><th></th><th>Binary Response</th><th>Survival Response</th></tr>
<tr><td rowspan="5">model is unspecified</td><td>$\hat{\gamma}$ (MLE) converges to γ^*</td><td>$\hat{\gamma}$ (partial MLE) converges to γ^*</td></tr>
<tr><td colspan="2">$W(z)'\gamma^*/2$ can be solved by</td></tr>
<tr><td>$$\max_f \mathrm{E}\left\{Yf(\mathbf{Z})T - \log(1 + e^{f(\mathbf{Z})T})\right\}$$ subject to $f \in \mathcal{F} = \{\gamma'\mathbf{W}(\mathbf{z})/2 \mid \boldsymbol{\gamma} \in R^p\}$,</td><td>$$\max_f \mathrm{E} \int_0^\tau \left(f(\mathbf{Z})T - \log[\mathrm{E}\{e^{f(\mathbf{Z})T} I(X \geq u)\}]\right) dN(u)$$ subject to $f \in \mathcal{F} = \{\boldsymbol{\gamma}'\mathbf{W}(\mathbf{z})/2 \mid \boldsymbol{\gamma} \in R^p\}$,</td></tr>
<tr><td colspan="2">the minimizer f*</td></tr>
<tr><td>$$f^*(\mathbf{z}) = \log\left\{\frac{1 - \Delta(\mathbf{z})}{1 + \Delta(\mathbf{z})}\right\}$$</td><td>$$e^{f^*(\mathbf{z})}\mathrm{E}\{\Lambda^*(\tilde{X}^{(1)}) \mid \mathbf{Z} = \mathbf{z}\} - e^{-f^*(\mathbf{z})}\mathrm{E}\{\Lambda^*(\tilde{X}^{(-1)}) \mid \mathbf{Z} = \mathbf{z}\}$$ $$= \mathrm{P}(\delta = 1 \mid T = 1, \mathbf{Z} = \mathbf{z}) - \mathrm{P}(\delta = 1 \mid T = -1, \mathbf{Z} = \mathbf{z})$$</td></tr>
</table>

Regularization for the High-Dimensional Data

- Method: variable selection procedures
- Eg. L_1 penalized estimator (LASSO) γ can be estimated by minimizing

$$\frac{1}{N}\sum_{i=1}^{N} l(Y_i, \boldsymbol{\gamma}'\mathbf{W}_i^*) + \lambda_N \|\boldsymbol{\gamma}\|_1, \qquad (9)$$

where $\|\boldsymbol{\gamma}\|_1 = \sum_{j=1}^{p} |\gamma_j|$ and

$$l(Y_i, \boldsymbol{\gamma}'\mathbf{W}_i^*) = \begin{cases} \frac{1}{2}(Y_i - \boldsymbol{\gamma}'\mathbf{W}_i^*)^2 \\ \qquad \text{for continuous responses} \\ -\{Y_i\boldsymbol{\gamma}'\mathbf{W}_i^* - \log(1 + e^{\boldsymbol{\gamma}'\mathbf{W}_i^*})\} \\ \qquad \text{for binary responses} \\ -\left[\boldsymbol{\gamma}'\mathbf{W}_i^* - \log\{\sum_{j=1}^{N} e^{\boldsymbol{\gamma}'\mathbf{W}_j^*} I(X_j \geq X_i)\}\right]\delta_i \\ \qquad \text{for survival responses.} \end{cases}$$

Efficiency Augment

- The estimator $\hat{\gamma}$ is

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmin}_{\boldsymbol{\gamma}} \frac{1}{N} \sum_{i=1}^{N} l(Y_i, \boldsymbol{\gamma}' \mathbf{W}_i^*).$$

- Using a nonrandom function $\boldsymbol{a}(\boldsymbol{z})$: $R^p \rightarrow R^q$, $E\{T_i a(Z_i)\} = 0$ and

$$\hat{\boldsymbol{\gamma}} = \operatorname{argmin}_{\boldsymbol{\gamma}} \frac{1}{N} \sum_{i=1}^{N} \left\{ l(Y_i, \boldsymbol{\gamma}' \mathbf{W}_i^*) - T_i \mathbf{a}(\mathbf{Z}_i)' \boldsymbol{\gamma} \right\}$$

- Optimal choice for continuous repsonses is $\mathbf{a}_0(\mathbf{z}) = -\frac{1}{2} \mathbf{W}(\mathbf{z}) \mathrm{E}(Y | \mathbf{Z} = \mathbf{z})$
- Optimal choice for binary responses is $\mathbf{a}_0(\mathbf{z}) = -\frac{1}{2} \mathbf{W}(\mathbf{z}) \{\mathrm{E}(Y | \mathbf{Z} = \mathbf{z}) - 0.5\}$

Augmented Estimator γ^*

- We can follow two-step procedures to estimate γ^*:
- 1. Estimate the optimal $\boldsymbol{a}_0(\boldsymbol{z})$:
  - (a) continuous responses (using linear model):

$$\boldsymbol{a}_0(\boldsymbol{z}) = -\frac{1}{2}\boldsymbol{W}(\boldsymbol{z})E(Y|\boldsymbol{Z}=\boldsymbol{z}) = -\frac{1}{2}\boldsymbol{W}(\boldsymbol{Z})\hat{\xi}'\boldsymbol{B}(\boldsymbol{z})$$

  - (b) binary response (using logit model):

$$\boldsymbol{a}_0(\boldsymbol{z}) = -\frac{1}{2}\boldsymbol{W}(\boldsymbol{z})\left(E(Y|\boldsymbol{Z}=\boldsymbol{z}) - \frac{1}{2}\right) = -\frac{1}{2}\boldsymbol{W}(\boldsymbol{z})\left(\frac{e^{\hat{\xi}'\boldsymbol{B}(\boldsymbol{z})}}{1+e^{\hat{\xi}'\boldsymbol{B}(\boldsymbol{z})}} - \frac{1}{2}\right)$$

Augmented Estimator γ^*

- 2. Estimate γ^*
  - (a) continuous response:
  - $\hat{\gamma}^* = \operatorname{argmin}_\gamma \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{2}(Y_i - \boldsymbol{\gamma}'\mathbf{W}_i^*)^2 - \boldsymbol{\gamma}'\hat{\mathbf{a}}(\mathbf{Z}_i)T_i \right\}$

    $= \operatorname{argmin}_\gamma \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left\{ Y_i - \hat{\xi}'\mathbf{B}(\mathbf{Z}_i) - \frac{1}{2}\boldsymbol{\gamma}'\mathbf{W}(\mathbf{Z}_i)T_i \right\}^2$
  - (b) binary response:
  - $\hat{\gamma}^* = \operatorname{argmin}_\gamma \frac{1}{N} \sum_{i=1}^{N} \left[-\{Y_i \boldsymbol{\gamma}'\mathbf{W}_i^* - \log(1 + e^{\boldsymbol{\gamma}'\mathbf{W}_i^*})\} - \boldsymbol{\gamma}'\hat{\mathbf{a}}(\mathbf{Z}_i)T_i \right]$

Survival outcome situation

- The optimal choice of $a(z)$ is

$$\mathbf{a}_0(\mathbf{z}) = -\frac{1}{2}\left[\frac{1}{2}\mathbf{W}(\mathbf{z})\left\{G_1(\tau;\mathbf{z}) + G_2(\tau;\mathbf{z})\right\} - \int_0^{\tau} \mathbf{R}(u;\boldsymbol{\gamma}^*)\{G_1(du;\mathbf{z}) - G_2(du;\mathbf{z})\}\right]$$

where $G_1(u;\mathbf{z}) = E\{M(u)|\mathbf{Z} = \mathbf{z}, T = 1\}, G_2(u;\mathbf{z}) = E\{M(u)|\mathbf{Z} = \mathbf{z}, T = -1\}$,

$$M(t, \mathbf{W}^*, \boldsymbol{\gamma}) = N(t) - \int_0^t \frac{I(X \geq u)e^{\boldsymbol{\gamma}'\mathbf{W}^*} d\mathrm{E}\{N(u)\}}{\mathrm{E}\{e^{\boldsymbol{\gamma}'\mathbf{W}^*} I(X \geq u)\}}$$

and

$$\mathbf{R}(u;\boldsymbol{\gamma}) = \frac{\mathrm{E}\{\mathbf{W}^* e^{\boldsymbol{\gamma}'\mathbf{W}^*} I(X \geq u)\}}{\mathrm{E}\{e^{\boldsymbol{\gamma}'\mathbf{W}^*} I(X \geq u)\}}.$$

Survival outcome situation

- For the high-dimensional case, the interaction effect is small so that we can assume $\gamma^* \approx 0$.
- If the sensoring patterns are similar in both groups, we also can assume $G_1(u, \mathbf{z}) \approx G_2(u, \mathbf{z})$.
- Thus, we can simplify the optimal choice as

$$\mathbf{a}_0(\mathbf{z}) = -\frac{1}{4}\mathbf{W}(\mathbf{z})\{G_1(\tau;\mathbf{z}) + G_2(\tau;\mathbf{z})\} = -\frac{1}{2}\mathbf{W}(\mathbf{z}) \times \mathrm{E}\{M(\tau)|\mathbf{Z} = \mathbf{z}),$$

where

$$M(t) = N(t) - \int_0^t \frac{I(X \geq u)d\mathrm{E}\{N(u)\}}{\mathrm{E}\{I(X \geq u)\}}.$$

Augmented Estimator γ^*

- Similarly, we can follow two-step procedures to estimate γ^*:
- 1. Estimate the optimal $\boldsymbol{a}_0(\boldsymbol{z})$:
  - (i) Calculate $\widehat{M_i}(\boldsymbol{\tau})$

$$\hat{M}_i(\tau) = N_i(\tau) - \int_0^\tau \frac{I(X_i \geq u) d\{\sum_{j=1}^N N_j(u)\}}{\sum_{j=1}^N I(X_j \geq u)}, \; i = 1, \ldots, N$$

  - (ii) Fit the linear regression model: $\mathrm{E}(\hat{M}(\tau)|\mathbf{Z}) = \xi'\mathbf{B}(\mathbf{Z})$
  - (iii) Calculate $\widehat{\boldsymbol{a}}(\boldsymbol{z})$

$$\hat{\mathbf{a}}(\mathbf{z}) = -\frac{1}{2}\mathbf{W}(\mathbf{z}) \times \hat{\xi}'\mathbf{B}(\mathbf{z}).$$

Augmented Estimator γ^*

- 2. Estimate γ^*
- $\gamma^* = \text{argmin}_{\gamma} \; \frac{1}{N}\sum_{i=1}^{N}\left(-\left[\boldsymbol{\gamma}'\mathbf{W}_i^* - \log\left\{\sum_{j=1}^{N} e^{\boldsymbol{\gamma}'\mathbf{W}_i^*} I(X_j \geq X_i)\right\}\right] \times \Delta_i - \boldsymbol{\gamma}'\hat{\mathbf{a}}(\mathbf{Z}_i)T_i\right).$