Supplementary Materials for "A Simple Method for Estimating Interactions between a Treatment and a Large Number of Covariates"

LU TIAN *

ASH A ALIZADEH †

ANDREW J GENTLES ‡

and

ROBERT TIBSHIRANI§

1 Estimating the Personalized Treatment Effect in Terms of Relative Risk

When the response Y is binary, relative risk may also be used as a measure for individualized treatment effects. For example, if we consider an alternative approach for fitting the logistic

*Department of Health Research & Policy, Stanford University, Palo Alto, CA 94305. lutian@stanford.edu

†Department of Medicine, Stanford University, Palo Alto, CA 94305.

‡Integrative Cancer Biology Program, Stanford University, Palo Alto, CA 94305.

§Department of Health Research & Policy and Department of Statistics, Stanford University, Palo Alto, CA 94305.

regression working model with modified covariates $\mathbf{W}_i^* = T_i\mathbf{W}_i/2$ by letting

$$\hat{\boldsymbol{\gamma}} = \mathrm{argmin}_{\boldsymbol{\gamma}} \sum_{i=1}^{N} \left\{ (1 - Y_i)\boldsymbol{\gamma}'\mathbf{W}^* + Y_i e^{-\boldsymbol{\gamma}'\mathbf{W}_i^*} \right\},$$

then $\hat{\boldsymbol{\gamma}}$ converges to a deterministic limit $\tilde{\boldsymbol{\gamma}}^*$ and $\exp\{\mathbf{W}(\mathbf{z})'\tilde{\boldsymbol{\gamma}}^*(\mathbf{z})/2\}$ can be viewed as an approximation to

$$\widetilde{\Delta}(\mathbf{z}) = \frac{\mathrm{P}(Y^{(1)} = 1|\mathbf{Z} = \mathbf{z})}{\mathrm{P}(Y^{(-1)} = 1|\mathbf{Z} = \mathbf{z})}, \tag{1}$$

which measures the treatment effect using "relative risk" rather than "risk difference". This loss function is motivated by the fact that the logistic regression model can be fitted by solving the estimating equation

$$N^{-1} \sum_{i=1}^{N} \left[\mathbf{W}_i^* \left\{ (1 - Y_i) - Y_i e^{-\boldsymbol{\gamma}'\mathbf{W}_i^*} \right\} \right] = 0,$$

which is the derivative of the proposed loss function. Furthermore, the optimal augmentation term $a_0(\mathbf{z})$ for this estimating function can be approximated by

$$-\frac{1}{2}\mathbf{W}(\mathbf{z}) \left\{ \mathrm{E}(Y|\mathbf{Z} = \mathbf{z}) - \frac{1}{2} \right\}$$

when $\boldsymbol{\gamma}^* \approx 0$. The efficiency augmentation algorithm can be carried out accordingly.

To justify (1), we consider the proposed objective function

$$\tilde{l}(Y, f(\mathbf{Z})T) = (1 - Y)f(\mathbf{Z})T + Ye^{-f(\mathbf{Z})T},$$

and have

$$
\begin{aligned}
\mathcal{L}(f) =& \mathrm{E}\{\tilde{l}(Y, f(\mathbf{Z})T)\} \\
=& \mathrm{E}_{\mathbf{Z}}\left[\frac{1}{2}\mathrm{E}_Y\{l(Y, f(\mathbf{Z})T)|\mathbf{Z}, T=1\} + \frac{1}{2}\mathrm{E}_Y\{l(Y, f(\mathbf{Z})T)|\mathbf{Z}, T=-1\}\right] \\
=& \mathrm{E}_{\mathbf{Z}}\left[\frac{1}{2}\{m_{-1}(\mathbf{Z}) - m_1(\mathbf{Z})\}f(\mathbf{Z}) + \frac{1}{2}m_1(\mathbf{Z})e^{-f(\mathbf{Z})} + \frac{1}{2}m_{-1}(\mathbf{Z})e^{f(\mathbf{Z})}\right],
\end{aligned}
$$

where $m_t(\mathbf{z}) = \mathrm{P}(Y^{(t)} = 1|\mathbf{Z} = \mathbf{z})$ for $t = 1$ or -1. Therefore

$$
\frac{\partial \mathcal{L}(f)}{\partial f} = \frac{1}{2}\mathrm{E}_{\mathbf{Z}}\left[\{m_{-1}(\mathbf{Z}) - m_1(\mathbf{Z})\} - m_1(\mathbf{Z})e^{-f(\mathbf{Z})} + m_{-1}(\mathbf{Z})e^{f(\mathbf{Z})}\right]
$$

which implies that the minimizer of $\mathcal{L}(f)$ is

$$
f^*(\mathbf{z}) = \log \frac{m_1(\mathbf{z})}{m_{-1}(\mathbf{z})}
$$

for all $\mathbf{z} \in$ Support of $\mathbf{Z}$.

2 Justification of the Optimal $a_0(\mathbf{z})$ in the Efficiency Augmentation

Let $S(y, \mathbf{w}^*, \boldsymbol{\gamma})$ be the derivative of the objective function $l(y, \boldsymbol{\gamma}'\mathbf{w}^*)$ with respect to $\boldsymbol{\gamma}$. $\hat{\boldsymbol{\gamma}}$ is the root of an estimating equation

$$
Q(\boldsymbol{\gamma}) = N^{-1}\sum_{i=1}^{N} S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}) = 0.
$$

Similarly, the augmented estimator $\hat{\boldsymbol{\gamma}}_a$ can be viewed as the root of the estimating equation

$$Q_a(\boldsymbol{\gamma}) = N^{-1}\sum_{i=1}^{N}\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}) - T_i \cdot \mathbf{a}(\mathbf{Z}_i)\} = 0.$$

Since $E\{T_i \cdot \mathbf{a}(\mathbf{Z}_i)\} = 0$ due to randomization, the solution of the augmented estimating equation always converges to $\boldsymbol{\gamma}^*$ in probability. It is straightforward to show that

$$\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}^* = N^{-1}A_0^{-1}\sum_{i=1}^{N} S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}^*) + o_P(N^{-1})$$

and

$$\hat{\boldsymbol{\gamma}}_a - \boldsymbol{\gamma}^* = N^{-1}A_0^{-1}\sum_{i=1}^{N}\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}^*) - T_i\mathbf{a}(\mathbf{Z}_i)\} + o_P(N^{-1}),$$

where A_0 is the derivative of $\mathrm{E}\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma})\}$ with respect to $\boldsymbol{\gamma}$ at $\boldsymbol{\gamma} = \boldsymbol{\gamma}^*$. Selecting the optimal $\mathbf{a}(\mathbf{z})$ is equivalent to minimizing the variance of $\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}^*) - T_i\mathbf{a}(\mathbf{Z}_i)\}$. Noting that

$$\mathrm{E}\left[\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}^*) - T_i\mathbf{a}(\mathbf{Z}_i)\}^{\otimes 2}\right] = \mathrm{E}\left[\{S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}^*) - T_i\mathbf{a}_0(\mathbf{Z}_i)\}^{\otimes 2}\right] + \mathrm{E}[\{\mathbf{a}(\mathbf{Z}_i) - \mathbf{a}_0(\mathbf{Z}_i)\}^{\otimes 2}],$$

where $\mathbf{a}_0(\mathbf{z})$ satisfies the equation

$$E\left[\{S(Y, \mathbf{W}^*, \boldsymbol{\gamma}^*) - T\mathbf{a}_0(\mathbf{Z})\}T\eta(\mathbf{Z})\right] = 0$$

for any function $\eta(\cdot)$, $\mathbf{a}_0(\cdot)$ is the optimal augmentation term minimizing the variance of $\hat{\boldsymbol{\gamma}}_a$. Since $\mathbf{a}_0(\cdot)$ is the root of the equation

$$\mathrm{E}\left[\{S(Y, \mathbf{W}^*, \boldsymbol{\gamma}^*) - T\mathbf{a}_0(\mathbf{Z})\}'T \,\middle|\, \mathbf{Z} = \mathbf{z}\right] = 0,$$

$$\mathbf{a}_0(\mathbf{z}) = \frac{1}{2}\left[\mathrm{E}\{S(Y, \mathbf{W}(\mathbf{z})/2, \boldsymbol{\gamma}^*)|\mathbf{Z} = \mathbf{z}, T = 1\} - \mathrm{E}\{S(Y, -\mathbf{W}(\mathbf{z})/2, \boldsymbol{\gamma}^*)|\mathbf{Z} = \mathbf{z}, T = -1\}\right].$$

For continuous responses,

$$S(Y, \mathbf{W}^*, \boldsymbol{\gamma}) = -\frac{1}{2} T \mathbf{W}(\mathbf{Z}) \left\{ Y - \frac{1}{2} T \mathbf{W}(\mathbf{Z})' \boldsymbol{\gamma} \right\}$$

and

$$\begin{aligned} a_0(\mathbf{z}) =& \frac{1}{2} \left(\mathrm{E}[-\mathbf{W}(\mathbf{z})\{Y - \mathbf{W}(\mathbf{z})'\boldsymbol{\gamma}^*/2\}/2 | T = 1, \mathbf{Z} = \mathbf{z}] - \mathrm{E}[\mathbf{W}(\mathbf{z})\{Y + \mathbf{W}(\mathbf{z})'\boldsymbol{\gamma}^*/2\}/2 | T = -1, \mathbf{Z} = \mathbf{z}] \right) \\ =& - \mathbf{W}(\mathbf{z}) \left\{ \frac{1}{4} \mathrm{E}(Y | T = 1, \mathbf{Z} = \mathbf{z}) + \frac{1}{4} \mathrm{E}(Y | T = -1, \mathbf{Z} = \mathbf{z}) \right\} \\ =& - \frac{1}{2} \mathbf{W}(\mathbf{z}) \mathrm{E}(Y | \mathbf{Z} = \mathbf{z}). \end{aligned}$$

For binary responses,

$$S(Y, \mathbf{W}^*, \boldsymbol{\gamma}) = -\frac{1}{2} \mathbf{W}(\mathbf{Z}) T \left\{ Y - \frac{e^{T \mathbf{W}(\mathbf{Z})' \boldsymbol{\gamma}/2}}{1 + e^{T \mathbf{W}(\mathbf{Z})' \boldsymbol{\gamma}/2}} \right\}$$

and

$$\begin{aligned} a_0(\mathbf{z}) =& - \frac{1}{4} \mathbf{W}(\mathbf{z}) \left[\mathrm{E} \left\{ Y - \frac{e^{\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}}{1 + e^{\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}} \,\middle|\, T = 1, \mathbf{Z} = \mathbf{z} \right\} + \mathrm{E} \left\{ Y - \frac{e^{-\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}}{1 + e^{-\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}} \,\middle|\, T = -1, \mathbf{Z} = \mathbf{z} \right\} \right] \\ =& - \frac{1}{4} \mathbf{W}(\mathbf{z}) \left\{ \mathrm{E}(Y | T = 1, \mathbf{Z} = \mathbf{z}) + \mathrm{E}(Y | T = -1, \mathbf{Z} = \mathbf{z}) - \left(\frac{e^{\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}}{1 + e^{\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}} + \frac{e^{-\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}}{1 + e^{-\mathbf{W}(\mathbf{z})' \boldsymbol{\gamma}^*/2}} \right) \right\} \\ =& - \frac{1}{2} \mathbf{W}(\mathbf{z}) \left\{ \mathrm{E}(Y | \mathbf{Z} = \mathbf{z}) - \frac{1}{2} \right\}. \end{aligned}$$

For survival responses, the estimating equation based on the partial likelihood function is asymptotically equivalent to the estimating equation $N^{-1}\sum_{i=1}^{N} S(Y_i, \mathbf{W}_i^*, \boldsymbol{\gamma}) = 0$, where

$$S(Y, \mathbf{W}^*, \boldsymbol{\gamma}) = -\int_0^{\tau} [\mathbf{W}^* - \mathbf{R}(u; \boldsymbol{\gamma})]\, M(du, \mathbf{W}^*, \boldsymbol{\gamma}).$$

Thus

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

3 The Lasso Algorithm in the Efficiency Augmentation

In general, the augmentation term is in the form of $\mathbf{a}_0(\mathbf{Z}_i) = \mathbf{W}(\mathbf{Z}_i)'\hat{r}(\mathbf{Z}_i)$, where $\hat{r}(\mathbf{Z}_i)$ is a simple scalar. The Lasso regularized objective function can be written as

$$\frac{1}{N}\sum_{i=1}^{N}\{l(Y_i, \boldsymbol{\gamma}'\mathbf{W}_i^*) - \boldsymbol{\gamma}'\mathbf{W}_i^*\hat{r}(\mathbf{Z}_i)\} + \lambda|\boldsymbol{\gamma}|.$$

In general, this Lasso problem can be solved iteratively. For example, when $l(\cdot)$ is the negative log-likelihood function of the logistic regression model, we may update $\hat{\boldsymbol{\gamma}}$ iteratively by solving the standard OLS-Lasso problem

$$\frac{1}{N}\sum_{i=1}^{N}\hat{w}_i(\hat{z}_i - \boldsymbol{\gamma}'\mathbf{W}_i^*)^2 + \lambda\|\boldsymbol{\gamma}\|_1,$$

where

$$\hat{z}_i = \hat{\boldsymbol{\gamma}}'\mathbf{W}_i^* + \hat{w}_i^{-1}\{Y_i - \hat{p}_i + \hat{r}(\mathbf{Z}_i)\}, \qquad \hat{w}_i = \hat{p}_i(1 - \hat{p}_i),$$

$\hat{\boldsymbol{\gamma}}$ is the current estimator for $\boldsymbol{\gamma}$ and

$$\hat{p}_i = \frac{\exp\{\hat{\boldsymbol{\gamma}}'\mathbf{W}_i^*\}}{1 + \exp\{\hat{\boldsymbol{\gamma}}'\mathbf{W}_i^*\}}.$$