Supplementary material for “Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data”

BY WEIHUA CAO, ANASTASIOS A. TSIATIS, and MARIE DAVIDIAN

Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, U.S.A.

wcao5@ncsu.edu tsiatis@stat.ncsu.edu davidian@stat.ncsu.edu

In this note, we provide further information and details on material appearing in the paper “Improving efficiency and robustness of the doubly robust estimator for a population mean with incomplete data,” by Cao, Tsiatis and Davidian. In § A, we sketch how the approach to identifying the optimal, doubly robust estimator for a single treatment mean in § 3 of the main paper may be adapted to estimation of a difference in treatment means as in the case of a causal treatment effect. In § 2, we follow up on our comment that “the asymptotic variance of the estimator for \( \mu \) can be approximated by the usual empirical sandwich technique” (§ 3 of the main paper) and give the explicit form of the building blocks required to calculate the asymptotic variances of the four estimators considered in the paper: \( \hat{\mu}_{OR} \), \( \hat{\mu}_{USUAL} \), \( \hat{\mu}_{TAN} \) and \( \hat{\mu}_{PROJ} \). Throughout, we use the notation defined in the main paper.

A. OPTIMAL, DOUBLY ROBUST ESTIMATOR FOR A TREATMENT MEAN DIFFERENCE

We consider the case where the propensity score is fully specified to demonstrate the idea. Extension to the situation where the a parametric propensity score model is postulated and fitted by binary regression techniques would be handled analogously to the developments in § 3 of the main paper.
Suppose we observe independent and identically distributed data \((R_i, Y_i, X_i), i = 1, \ldots, n\), where now \(R_i\) is a treatment indicator taking values 0 or 1, and \(Y_i\) and \(X_i\) are response and covariates, respectively, as in the main paper. (Note that these observed data are different from those in the paper, \((R_i Y_i, Y_i, X_i)\).) As in the standard set-up for the usual causal inference problem in an observational point exposure study, we may conceptualize potential outcomes \(Y_0\) and \(Y_1\) corresponding to each treatment, with the observed \(Y = RY_1 + (1-R)Y_0\). The goal is to estimate the difference \(\Delta = E(Y_1) - E(Y_0)\) based on the observed data under the usual assumption of no unmeasured confounders, namely, that \((Y_0, Y_1)\) are independent of \(R\) given \(X\).

Consider as an estimator for \(\Delta\)

\[
n^{-1} \sum_{i=1}^{n} \left\{ \frac{R_i Y_i}{\pi(X_i)} - \frac{(1-R_i)Y_i}{1-\pi(X_i)} - \{R_i - \pi(X_i)\} m(X_i, \hat{\beta}) \right\}
\]

for some \(m(X, \beta)\) and estimator \(\hat{\beta}\) for \(\beta\). Following the ideas in § 3 of the main paper, we wish to identify a value \(\beta^*_{opt}\) and an estimator \(\hat{\beta}\) converging to it such that the asymptotic variance of (1) is minimized when the propensity score is correct and such that (1) is doubly robust. Using the assumption of no unmeasured confounders, it is straightforward to deduce that the optimal value \(\beta^*_{opt}\) is the solution in \(\beta\) to

\[
E \left[ \left\{ \frac{Y_1}{\pi_0(X)} + \frac{Y_0}{1-\pi_0(X)} - m(X, \beta) \right\} \pi_0(X) \{1 - \pi_0(X)\} m_\beta(X, \beta) \right] = 0,
\]

or equivalently

\[
E \left[ \left\{ \frac{m^{(1)}(X)}{\pi_0(X)} + \frac{m^{(0)}(X)}{1-\pi_0(X)} - m(X, \beta) \right\} \pi_0(X) \{1 - \pi_0(X)\} m_\beta(X, \beta) \right] = 0, \tag{2}
\]

where \(m^{(k)}(X) = E(Y_k \mid X), k = 0, 1,\) are the true potential outcome means conditional on \(X\). Suppose we posit models \(m_k(X, \alpha_k)\) for \(E(Y_k \mid X), k = 0, 1,\) where the models \(m_k(X, \alpha_k)\)
are possibly nonlinear in $\alpha_k$, and take

$$m(X,\beta) = \frac{m_0(X,\alpha_0)}{1-\pi(X)} + \frac{m_1(X,\alpha_1)}{\pi(X)}, \quad \beta = (\alpha_0^T, \alpha_1^T)^T.$$  

Suppose we estimate $\beta$ by solving the estimating equation

$$n^{-1} \sum_{i=1}^{n} \left[ \frac{R_i(1-\pi(X_i))}{\pi(X_i)} \{Y_i - m_1(X,\alpha_1)\} + \frac{(1-R_i)\pi(X_i)}{1-\pi(X_i)} \{Y_i - m_0(X,\alpha_0)\} \right] m(\beta) = 0$$  

(3)

If the propensity score is correct, $\pi(X) = \pi_0(X)$, but the models $m_k(X,\alpha_k), k = 0, 1$, are not, then it is straightforward to show that the left hand side of (3) converges to an expression of the form of the left hand side of (2), so that the estimator $\hat{\beta}$ solving (3) converges in probability to $\beta_{opt}^*$. On the other hand, if the propensity score model is not correct but the models $m_k(X,\alpha_k)$ are in the sense that there are values $\alpha_k^{(0)}$ such that $m_k(X,\alpha_k^{(0)}) = m^{(k)}(X)$, $k = 0, 1$, then the left hand side of (3) converges to

$$E \left[ \frac{\pi_0(X)}{1-\pi(X)} \{m^{(1)}(X) - m_1(X,\alpha_1)\} m(\beta) + \frac{1-\pi_0(X)}{1-\pi(X)} \{m^{(0)}(X) - m_0(X,\alpha_0)\} m(\beta) \right],$$

which equals zero when $\beta = \beta_0 = (\alpha_0^{(0)T}, \alpha_1^{(0)T})^T$, so that $\hat{\beta}$ converges in probability to $\beta_0$.

This shows that the estimator for $\Delta$ in (1) is doubly robust and achieves the minimum asymptotic variance even if the models $m_k(X,\alpha_k)$ for $E(Y_k \mid X), k = 0, 1$, are misspecified.

**B. Calculation of asymptotic variances**

We provide expressions required to calculate the asymptotic variances of the four estimators considered in the paper: $\hat{\mu}_{OR}$, $\hat{\mu}_{USUAL}$, $\hat{\mu}_{TAN}$ and $\hat{\mu}_{PROJ}$. The asymptotic variance of $\hat{\mu}_{USUAL}^{en}$ and $\hat{\mu}_{PROJ}^{en}$ may be calculated using the same formulæ as for $\hat{\mu}_{USUAL}$ and $\hat{\mu}_{PROJ}$; see below.
The following results are valid for a general propensity score model, including the logistic model or the enhanced model discussed in § 4 of the main paper. In order to streamline presentation of the results, denote the propensity score model by \( \pi_i = \pi(X_i, \gamma) \). Similarly, denote the outcome regression model by \( m_i = m(X_i, \beta) \). Let the score corresponding to the propensity score model be \( S_{\gamma i} = S_{\gamma}(R_i, X_i, \gamma) \), and define \( S_{\gamma \gamma i} = \partial / \partial \gamma^T S_{\gamma}(R_i, X_i, \gamma) \). Write \( \pi_{\gamma i} = \pi_{\gamma}(X_i, \gamma) \), \( m_{\beta i} = m_{\beta}(X_i, \beta) \), \( \pi_{\gamma \gamma i} = \partial^2 / \partial \gamma \partial \gamma^T \{ \pi(X_i, \gamma) \} \), and \( \pi_{\gamma \gamma i} = \partial^2 / \partial \gamma \partial \gamma^T \{ \pi(X_i, \gamma) \} \).

Let \( \xi \) be collection of unknown parameters involved in obtaining the estimators for \( \mu \); in particular, \( \xi = (\beta^T, \mu)^T \) for \( \hat{\mu}_{OR} \), \( \xi = (\gamma^T, \beta^T, \mu)^T \) for \( \hat{\mu}_{USUAL} \), \( \xi = (\gamma^T, \beta^T, c^T, \mu)^T \) for \( \hat{\mu}_{PROJ} \), and \( \xi = (\gamma^T, \beta^T, \alpha_0, \alpha_1, c^T, \mu)^T \) for \( \hat{\mu}_{TAN} \).

The estimator for \( \xi, \hat{\xi} \), in each case can be obtained by solving a set of M-estimating equations \( \sum_{i=1}^n \psi_i(\xi) = 0 \) (Stefanski & Boos, 2002), where the last element of \( \psi_i(\xi) \) corresponds to the estimating equation for \( \mu \). Let \( A_n = n^{-1} \sum_{i=1}^n A_i = n^{-1} \sum_{i=1}^n \partial / \partial \xi \{ \psi_i(\xi) \} \), and \( B_n = n^{-1} \sum_{i=1}^n \psi_i(\xi) \psi_i^T(\xi) \). Following standard theory, the asymptotic covariance matrix of \( \hat{\xi} \) can be approximated by the empirical sandwich matrix \( V_n = A_n^{-1} B_n (A_n^{-1})^T / n \). Therefore, the asymptotic variance of the four estimators can be approximated by the last, rightmost diagonal entry of the corresponding matrix \( V_n \). We present the form of \( \psi_i(\xi) \) and \( A_i \) for each of the estimators, from which the form of \( V_n \) may be calculated. The desired diagonal entry of \( V_n \) may then be obtained numerically, with the required matrix inversion carried out by standard routines.

For \( \hat{\mu}_{OR} \), \( \psi_i(\xi) \) is given by
\[
\psi_i(\xi) = \begin{pmatrix} m_{\beta i} R_i (Y_i - m_i) \\ m_i - \mu \end{pmatrix},
\]
and $A_i$ is given by

$$A_i = \begin{pmatrix} D_{1i} & 0 \\ m_{\beta i}^T & -1 \end{pmatrix}.$$ 

where $D_{1i} = m_{\beta i} R_i (Y_i - m_i) - R_i m_{\beta i} m_{\beta i}^T$.

For $\hat{\mu}_{\text{USUAL}}$, $\psi_i(\xi)$ is given by

$$\psi_i(\xi) = \begin{pmatrix} R_i - \pi_i \\ \pi_i (1 - \pi_i) \pi_{\gamma i} \\ m_{\beta i} R_i (Y_i - m_i) \\ \frac{R_i}{\pi_i} (Y_i - m_i) + m_i - \mu \end{pmatrix}. $$

$A_i$ is given by

$$A_i = \begin{pmatrix} D_{1i} & 0 & 0 \\ 0 & D_{2i} & 0 \\ D_{3i} & D_{4i} & -1 \end{pmatrix},$$

where $D_{1i} = \frac{R_i - \pi_i}{\pi_i (1 - \pi_i) \pi_{\gamma i}^2} (\pi_i (1 - \pi_i) + (R_i - \pi_i) (1 - 2\pi_i)) \pi_{\gamma i} \pi_{\gamma i}^T$, $D_{2i} = m_{\beta i} R_i (Y_i - m_i) - R_i m_{\beta i} m_{\beta i}^T$, $D_{3i} = -\frac{R_i}{\pi_i^2} (Y_i - m_i) \pi_{\gamma i}^T$, and $D_{4i} = \left(1 - \frac{R_i}{\pi_i}\right) m_{\beta i}^T$.

For $\hat{\mu}_{\text{TAN}}$, $\psi_i(\xi)$ is given by

$$\psi_i(\xi) = \begin{pmatrix} R_i - \pi_i \\ \frac{R_i}{\pi_i^2} (Y_i - m_i - \alpha_0 - \alpha_1 m_i - c^T \pi_{\gamma i}) \\ m_{\beta i} R_i (Y_i - m_i) \\ \frac{R_i}{\pi_i^2} m_i (Y_i - m_i - \alpha_0 - \alpha_1 m_i - c^T \pi_{\gamma i}) \\ \frac{R_i}{\pi_i^2} \frac{\pi_{\gamma i}}{1 - \pi_i} (Y_i - m_i - \alpha_0 - \alpha_1 m_i - c^T \pi_{\gamma i}) \\ \frac{R_i}{\pi_i} (Y_i - m_i - \alpha_0 - \alpha_1 m_i - c^T S_{\gamma i}) - \mu \end{pmatrix}. $$
\( A_i \) is given by

\[
\begin{pmatrix}
D_{1i} & 0 & 0 & 0 & 0 & 0 \\
0 & D_{2i} & 0 & 0 & 0 & 0 \\
D_{3i} & D_{4i} & D_{5i} & m_i D_{5i} & D_{6i} & 0 \\
m_i D_{3i} & 2m_i D_{4i} & m_i D_{5i} & m_i^2 D_5 & m_i D_6 & 0 \\
D_{7i} & t D_{4i} & t_i D_5 & t_i m_i D_5 & t_i D_{6i} & 0 \\
D_{8i} & D_{9i} & D_{10i} & m_i D_{10i} & -S_{\gamma i} & -1 \\
\end{pmatrix},
\]

where \( D_{1i} \) and \( D_{2i} \) are the same as for \( \hat{\mu}_{\text{USUAL}} \), \( D_{3i} = D_{31i} - c^T D_{32i} \),

\[
D_{31i} = \frac{-R_i \pi_i^2}{\pi_i^2} - 2 \frac{R_i (1 - \pi_i)}{\pi_i^2} \pi_i T \left( Y_i - \alpha_0 - \alpha_1 m_i - c^T \frac{\pi_{\gamma i}}{1 - \pi_i} \right),
\]

\[
D_{32i} = \frac{1}{(1 - \pi_i)^2} \frac{R_i (1 - \pi_i)}{\pi_i^2} \left( \frac{\pi_{\gamma i}}{1 - \pi_i} + \pi_{\gamma gamma i} \pi_i T \right),
\]

\[
D_{4i} = -\alpha_1 \frac{R_i (1 - \pi_i)}{\pi_i^2} m_{\beta i}, \quad D_{5i} = -\frac{R_i (1 - \pi_i)}{\pi_i^2} \pi_{\gamma i}, \quad D_{6i} = -\frac{R_i (1 - \pi_i)}{\pi_i^2} \pi_{\gamma i},
\]

\[
D_{7i} = \frac{\pi_{\gamma i}}{1 - \pi_i} D_{3i} + D_{32i}, \quad t_i = \frac{\pi_{\gamma i}}{1 - \pi_i}, \quad D_{8i} = -\frac{R_i}{\pi_i^2} \pi_{\gamma i} (Y_i - \alpha_0 - \alpha_1 m_i) - c^T S_{\gamma i},
\]

\[
D_{9i} = -\alpha_1 \frac{R_i}{\pi_i} m_{\beta i}^*, \quad \text{and} \quad D_{10i} = \frac{-R_i}{\pi_i} \pi_{\gamma i}.
\]

For \( \hat{\mu}_{\text{PROJ}} \), \( \psi_i(\xi) \) is given by

\[
\psi_i(\xi) = \begin{pmatrix}
\frac{R_i - \pi_i}{\pi_i (1 - \pi_i)} \pi_{\gamma i} \\
\frac{R_i (R_i - \pi_i)}{\pi_i^2} m_{\beta i} \left( Y_i - m_i - c^T \frac{\pi_{\gamma i}}{1 - \pi_i} \right) \\
\frac{R_i (R_i - \pi_i)}{\pi_i^2} \pi_{\gamma i} \left( Y_i - m_i - c^T \frac{\pi_{\gamma i}}{1 - \pi_i} \right) \\
\frac{R_i}{\pi_i} (Y_i - m_i) + m_i - c^T S_{\gamma i} - \mu
\end{pmatrix}.
\]
$A_i$ is given by

$$A_i = \begin{pmatrix}
D_{1i} & 0 & 0 & 0 \\
\frac{\partial m_i}{\partial \beta} D_{2i} & D_{3i} & D_{4i} & 0 \\
D_{5i} & D_{4i}^T & D_{6i} & 0 \\
D_{7i} & D_{8i} & -S_{\gamma i} & -1
\end{pmatrix},$$

where $D_{1i}$ is the same as that for $\hat{\mu}_{\text{TAN}}$, $D_{2i} = D_{21i} - c^T D_{22i}$,

$$D_{21i} = -R_i \pi^2_i - 2R_i(1 - \pi_i) \pi_{\gamma i} Y_i - m_i - c^T \pi_{\gamma i} \frac{1}{1 - \pi_i},$$

$$D_{22i} = \frac{1}{(1 - \pi_i)^2} \frac{R_i(1 - \pi_i)}{\pi_i^2} \left( \pi_{\gamma i} (1 - \pi_i) + \pi_{\gamma i} \pi_{\gamma i}^T \right),$$

$$D_{3i} = \frac{R_i(1 - \pi_i)}{\pi_i^2} m_{\beta \gamma i} \left( Y_i - m_i - c^T \pi_{\gamma i} \frac{1}{1 - \pi_i} \right) - \frac{R_i(1 - \pi_i)}{\pi_i^2} m_{\beta i} m_{\beta i}^T,$$

$$D_{4i} = -\frac{R_i(1 - \pi_i)}{\pi_i^2} m_{\beta i} \pi_{\gamma i} Y_i - m_i - c^T \pi_{\gamma i} \frac{1}{1 - \pi_i},$$

$$D_{5i} = \frac{\pi_{\gamma i}}{1 - \pi_i} D_{2i} + D_{22i},$$

$$D_{6i} = -\frac{1}{(1 - \pi_i)^2} \frac{R_i(1 - \pi_i)}{\pi_i^2} \pi_{\gamma i} \pi_{\gamma i}^T, D_{7i} = -\frac{R_i}{\pi_i^2} \pi_{\gamma i}^T (Y_i - m_i) - c^T S_{\gamma i},$$

and $D_{8i} = -\frac{R_i - \pi_i}{\pi_i} m_{\beta i}^T$.

In the main paper (§ 3), we also constructed an estimator by replacing $m(X, \beta)$ in $\hat{\mu}_{\text{PROJ}}$ by $\tilde{m}(X, \tilde{\beta}) = \alpha_0 + \alpha_1 m(X, \beta)$, $\tilde{\beta} = (\alpha_0, \alpha_1, \beta^T)^T$ and estimating all elements of $\tilde{\beta}$ simultaneously by solving the estimating equations (17). The asymptotic variance of the resulting estimator for $\mu$ can be calculated using the same formulæ for $\hat{\mu}_{\text{PROJ}}$ with $m(X, \beta)$ and $m_{\beta}(X, \beta)$ in $\psi_i(\xi)$ and $A_i$ replaced by $\tilde{m}(X, \tilde{\beta})$ and $\partial/\partial \tilde{\beta}\{\tilde{m}(X, \tilde{\beta})\}$, respectively.