Supplementary Material for Adaptive Response-Dependent Two-Phase Designs, by Michael A. McIsaac and Richard J. Cook

Asymptotic Variance of the Mean Score Estimator

Consider the weighted unbiased score equation

$$\overline{U}(\beta) = \sum_{i=1}^{N} \overline{U}_i(\beta) = \sum_{i=1}^{N} R_i \pi_i(Y_i, V_i)^{-1} U_\beta(Y_i|X_i, V_i) = 0$$

where we use $\pi_i = \pi(y_i, v_i) = P(R_i = 1|y_i, v_i) = n_{y_i v_i}/N_{y_i v_i}$ to denote the sampling probability for an individual i. As noted by Lawless et al. [1], under mild regularity conditions [2]

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{p} N(0, \mathcal{A}(\Psi)^{-1}\mathcal{C}(\Omega)\mathcal{A}(\Psi)^{-1}),$$

where $\Psi = (\beta', \alpha', \gamma')'$, $\Omega = (\Psi', \delta')'$,

$$\mathcal{A}(\Psi) = \lim E[-N^{-1}\partial\overline{U}_i(\beta)/\partial\beta'] = E_{YXV}[-\partial U_\beta(Y_i|X_i, V_i)/\partial\beta'],$$

and

$$\begin{aligned}
\mathcal{C}(\Omega) &= \lim\ \mathrm{var}(N^{-1/2}\sum_{i=1}^{N} R_i\pi_i^{-1}U_\beta(Y_i|X_i, V_i)) \\
&= \lim \Big\{ \mathrm{var}_{YXV}(E_{R|YXV}[N^{-1/2}\sum_{i=1}^{N} R_i\pi_i^{-1}U_\beta(Y_i|X_i, V_i)]) \\
&\quad + E_{YXV}(\mathrm{var}_{R|YXV}[N^{-1/2}\sum_{i=1}^{N} R_i\pi_i^{-1}U_\beta(Y_i|X_i, V_i)]) \Big\} \\
&= \lim\ N^{-1}\left\{ \mathrm{var}_{YXV}(\sum_{i=1}^{N} U_\beta(Y_i|X_i, V_i)) + E_{YXV}(\mathrm{var}_{R|YXV}[\sum_{i=1}^{N} R_i\pi_i^{-1}U_\beta(Y_i|X_i, V_i)]) \right\}.
\end{aligned}$$

We denote the second-order inclusion probability as $\pi_{ij} = P(R_i = 1, R_j = 1|y_i, v_i, y_j, v_j)$ so $\pi_{ij} = \pi_i \cdot \pi_j = n_{y_i v_i}/N_{y_i v_i} \cdot n_{y_j v_j}/N_{y_j v_j}$ if individuals i and j are from different strata (i.e. if $(y_i, v_i) \neq (y_j, v_j)$), while $\pi_{ij} = n_{y_i v_i}/N_{y_i v_i} \cdot (n_{y_i v_i} - 1)/(N_{y_i v_i} - 1)$ if they are from the same stratum (i.e. $(y_i, v_i) = (y_j, v_j)$).

So,

$$\begin{aligned}
&E_{YXV}\left[\mathrm{var}_{R|YXV}[\sum_{i=1}^{N} R_i\pi_i^{-1}U_\beta(Y_i|X_i, V_i)]\right] \\
&= E_{YXV}\Big[\sum_{i=1}^{N} \mathrm{var}_{R|YXV}(R_i)\pi_i^{-2}U_\beta(Y_i|X_i, V_i)U'_\beta(Y_i|X_i, V_i) + \\
&\qquad \sum_{i=1}^{N}\sum_{j=1;i\neq j}^{N} cov(R_i, R_j|Y_i, X_i, V_i, Y_j, X_j, V_j)\pi_i^{-1}\pi_j^{-1}U_\beta(Y_i|X_i, V_i)U'_\beta(Y_j|X_j, V_j)\Big] \\
&= E_{YXV}\Big[\sum_{i=1}^{N}(\pi_i - \pi_i^2)\pi_i^{-2}U_\beta(Y_i|X_i, V_i)U'_\beta(Y_i|X_i, V_i) + \sum_{i=1}^{N}\sum_{j=1;i\neq j}^{N}\frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}U_\beta(Y_i|X_i, V_i)U'_\beta(Y_j|X_j, V_j)\Big] \\
&= E_{YXV}\Big[\sum_{i=1}^{N}(\pi_i^{-1} - 1)U_\beta(Y_i|X_i, V_i)U'_\beta(Y_i|X_i, V_i)\Big] + \\
&\qquad \sum_{i=1}^{N} E_{YV}\left[(N_{YV} - 1)\left(\frac{n_{YV} - 1}{N_{YV} - 1}\frac{N_{YV}}{n_{YV}} - 1\right)E_{X|YV}[U_\beta(Y_i|X_i, V_i)]E_{X|YV}[U'_\beta(Y_i|X_i, V_i)]\right]
\end{aligned}$$

$$= E_{YXV}\big[\sum_{i=1}^{N}(\pi_i^{-1}-1)U_\beta(Y_i|X_i,V_i)U'_\beta(Y_i|X_i,V_i)\big] - \sum_{i=1}^{N} E_{YV}\left[\left(\pi_i^{-1}-1\right) E_{X|YV}[U_\beta(Y_i|X_i,V_i)]E_{X|YV}[U'_\beta(Y_i|X_i,V_i)]\right]$$

$$= \sum_{i=1}^{N} E_{YV}\left[(\pi_i^{-1}-1)\left(E_{X|YV}[U_\beta(Y_i|X_i,V_i)U'_\beta(Y_i|X_i,V_i)] - E_{X|YV}[U_\beta(Y_i|X_i,V_i)]E_{X|YV}[U'_\beta(Y_i|X_i,V_i)]\right)\right].$$

Therefore,

$$\mathcal{C}(\Omega) = E[U_\beta(Y_i|X_i,V_i)U'_\beta(Y_i|X_i,V_i)] + \sum_{YV} P(Y,V)(\pi(Y,V)^{-1}-1)\mathrm{var}_{X|YV}[U_\beta(Y_i|X_i,V_i)],$$

and since $E[U_\beta(Y_i|X_i,V_i)U'_\beta(Y_i|X_i,V_i)] = E_{YXV}[-\partial U_\beta(Y_i|X_i,V_i)/\partial\beta']$ [3], the asymptotic variance of the mean score estimator is

$$\mathcal{A}(\Psi)^{-1} + \mathcal{A}(\Psi)^{-1}\mathcal{B}(\Omega)\,\mathcal{A}(\Psi)^{-1}, \tag{1}$$

where

$$\mathcal{B}(\Omega) = \sum_{YV} P(Y,V)\big[\frac{N_{YV}}{n_{YV}} - 1\big]\cdot \mathrm{var}_{X|Y,V}[U_\beta(Y_i|X_i,V_i)].$$

References

1. Lawless JF, Kalbfleisch JD, Wild CJ. Semiparametric methods for response-selective and missing data problems in regression. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 1999; **61**(2):413–438.
2. Wild CJ. Fitting prospective regression models to case-control data. *Biometrika* 1991; **78**:705–717.
3. Pierce DA. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics* 1982; **10**:475–478.