

Asymptotic bias calculation for inefficient g-estimation on two intervals

More efficient g-estimation is achieved via

$$U_j(\psi) = (H_j(\psi) - E[H_j(\psi)|\bar{L}_j, \bar{A}_{j-1}])\{S_j(A_j) - E[S_j(A_j)|\bar{L}_j, \bar{A}_{j-1}]\}. \quad (1)$$

A less efficient, singly-robust version of equation (1) simply omits the expected counterfactual model:

$$U_j^*(\psi) = H_j(\psi)\{S_j(A_j) - E[S_j(A_j)|\bar{L}_j, \bar{A}_{j-1}]\}. \quad (2)$$

It is useful in the ensuing discussion to express the blip as $\gamma_j(\bar{L}_j, \bar{A}_j; \psi_j) = A_j g_j(\bar{L}_j, \bar{A}_{j-1}; \psi_j)$. Using this form, the optimal rule – and the definition of exceptional laws – can be written in terms of the function $g_j(\bar{L}_j, \bar{A}_{j-1}; \psi_j)$. For binary treatment, the optimal rule is defined by the indicator $I[g_j(\bar{L}_j, \bar{A}_{j-1}; \psi_j) > 0]$. Exceptional laws exist when $P(g_j(\bar{L}_j, \bar{A}_{j-1}; \psi_j) = 0) > 0$ and $g_j(\bar{L}_j, \bar{A}_{j-1}; \psi_j)$ depends on at least one component of $\bar{L}_j, \bar{A}_{j-1}$. Robins (2004, p. 315) calculates the asymptotic bias of $\hat{\psi}_1(\psi_2)$ found using equation (2) for linear blips in the special case where exceptional laws occur when $L_j = A_{j1} = 1$. We generalize the result here, and provide the asymptotic bias for (1) assuming a simple linear form for the expected counterfactual model in Appendix A:

Suppose a linear blip where $g_2(\bar{L}_2, A_1; \psi_2)$ depends on at least one of L_1, L_2, A_1 so that $g_2(\bar{L}_2, A_1; \psi_2) = 0$ whenever $\sum_{j=0}^{|\psi_2|-1} \psi_{2j} \cdot X_2 = 0$ where $|\psi_2|$ is the dimension of the vector ψ_2 and X_2 denotes the matrix of covariates (in our simulations, $X_2 = (1, L_2, A_1, L_2 A_1)$). Let $\psi_1^\dagger, \psi_2^\dagger$ denote the true values of ψ_1 and ψ_2 and let $\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) = U_1^*(\psi_1, \hat{\psi}_2) - U_1^*(\psi_1, \psi_2^\dagger)$. As indicated by the notation, $\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger)$ does not depend on ψ_1 when blips are linear. For

linear blip functions,

$$\begin{aligned}
\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) &= U_1(\psi_1, \hat{\psi}_2) - U_1(\psi_1, \psi_2^\dagger) \\
&= I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0] \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \\
&\quad \left(I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \hat{\psi}_2) \\
&\quad + I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) \neq 0] \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \\
&\quad \left\{ \left(I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \hat{\psi}_2) - \right. \\
&\quad \left. \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \psi_2^\dagger) \right\} \quad (3)
\end{aligned}$$

To find the asymptotic bias of $\hat{\psi}_1$, we perform a Taylor expansion, which gives

$$\begin{aligned}
(\hat{\psi}_1 - \psi_1^\dagger) &= - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} \mathbb{P}_n U_1^*(\psi_1^\dagger, \hat{\psi}_2) + o_p(n^{-1/2}) \\
&= - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} \mathbb{P}_n \left(U_1^*(\psi_1^\dagger, \psi_2^\dagger) + \Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right) + o_p(n^{-1/2}).
\end{aligned}$$

The asymptotic bias is equal to

$$\begin{aligned}
E[\hat{\psi}_1 - \psi_1^\dagger] &= - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} E \left[\mathbb{P}_n \left(U_1^*(\psi_1^\dagger, \psi_2^\dagger) + \Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right) \right] \\
&= - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} E \left[\mathbb{P}_n \Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right] \\
&= - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right],
\end{aligned}$$

since the g-estimating function $U_1^*(\psi_1^\dagger, \psi_2^\dagger)$ that contains the true values $\psi_1^\dagger, \psi_2^\dagger$ has mean zero.

The last equality follows from considering the difference term $\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger)$ as a function of random variables (rather than looking at observed values of covariates and responses).

Therefore, we must examine $E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right]$.

To proceed, we require an additional piece of information. We know that $\hat{\psi}_2$ is regular,

and so a Taylor expansion gives

$$\begin{aligned}(\hat{\psi}_2 - \psi_2^\dagger) &= -\mathbb{P}_n \left(\left(E \left[\frac{\partial}{\partial \psi_2} U_2^*(\psi_2^\dagger) \right] \right)^{-1} U_2^*(\psi_2^\dagger) \right) + o_p(n^{-1/2}) \\ &= \mathbb{P}_n (IF_{20}, IF_{21}, \dots, IF_{2, |\psi_2| - 1})^T + o_p(n^{-1/2}) = \mathbb{P}_n IF_2 + o_p(n^{-1/2}),\end{aligned}\quad (4)$$

where IF_{2j} denotes $-\left(E \left[\frac{\partial}{\partial \psi_{2j}} U_2^*(\psi_2^\dagger) \right]\right)^{-1} U_2^*(\psi_2^\dagger)$ and IF_2 is the vector of such functions.

To simplify the expressions we consider, we will examine $E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right]$ under non-exceptional laws and exceptional laws separately.

Calculations under non-exceptional laws:

Under non-exceptional laws, $g_2(\bar{L}_2, A_1; \psi_2^\dagger) \neq 0$ and

$$\begin{aligned}\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) &= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \left\{ \left(I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \hat{\psi}_2) \right. \\ &\quad \left. - \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \psi_2^\dagger) \right\}\end{aligned}\quad (5)$$

$$\begin{aligned}&= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \left\{ \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \hat{\psi}_2) \right. \\ &\quad \left. - \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \psi_2^\dagger) \right\}\end{aligned}\quad (6)$$

$$\begin{aligned}&= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) \cdot \\ &\quad \left(g_2(\bar{L}_2, A_1; \hat{\psi}_2) - g_2(\bar{L}_2, A_1; \psi_2^\dagger) \right)\end{aligned}\quad (7)$$

$$= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) (\hat{\psi}_2 - \psi_2^\dagger) \cdot X_2 \quad (8)$$

$$= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \cdot \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) (\mathbb{P}_n IF_2 \cdot X_2) + o_p(n^{-1/2})$$

Briefly, the steps above are justified by

(5)→(6) Since $I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) > 0]$ takes on the same sign as $I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0]$ with probability approaching 1 (Robins, 2004), we can substitute the latter for the former.

(6)→(7) Collecting terms.

(7)→(8) When blips are linear, $g_2(\bar{L}_2, A_1; \hat{\psi}_2) - g_2(\bar{L}_2, A_1; \psi_2^\dagger)$ can be expressed as $\hat{\psi}_2 \cdot X_2 - \psi_2^\dagger \cdot X_2$, where X_2 is a vector of 1's and some or all of L_1, A_1, L_2 such that $g_2(\bar{L}_2, A_1; \psi_2) = X_2' \psi_2$.

(8)→ Substitution using equation (4).

Now, $\bar{L}_2, A_1$ is ancillary for $\hat{\psi}_1$, so we may break the expectation of $\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger)$ into a product of expectations as follows:

$$E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right] = E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} \left(I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) \right] E[(\mathbb{P}_n IF_2 \cdot X_2)].$$

From this, we conclude that $E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right] = 0$ since $IF_2 \cdot X_2$ has mean zero and therefore that there is no asymptotic bias when laws are not exceptional.

Calculations under exceptional laws:

Under exceptional laws, $g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0$ and

$$\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) = \{S_1(A_1) - E[S_1(A_1)|L_1]\} \left(I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) > 0] - A_2 \right) g_2(\bar{L}_2, A_1; \hat{\psi}_2) \quad (9)$$

$$= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \left(I[g_2(\bar{L}_2, A_1; \hat{\psi}_2) - g_2(\bar{L}_2, A_1; \psi_2^\dagger) > 0] - A_2 \right) \cdot \left(g_2(\bar{L}_2, A_1; \hat{\psi}_2) - g_2(\bar{L}_2, A_1; \psi_2^\dagger) \right) \quad (10)$$

$$= \{S_1(A_1) - E[S_1(A_1)|L_1]\} \left(I[(\hat{\psi}_2 - \psi_2^\dagger) \cdot X_2 > 0] - A_2 \right) \left((\hat{\psi}_2 - \psi_2^\dagger) \cdot X_2 \right) \quad (11)$$

$$= \{S_1(A_1) - E[S_1(A_1)|L_1]\} (I[\mathbb{P}_n IF_2 \cdot X_2 > 0] - A_2) (\mathbb{P}_n IF_2 \cdot X_2)$$

The above steps are justified by

(9) → (10) Since the law is exceptional, $g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0$, and so this zero can be added into the equation, both inside the indicator function and at the end of the equation.

(10) → (11) When blips are linear, $g_2(\bar{L}_2, A_1; \hat{\psi}_2) - g_2(\bar{L}_2, A_1; \psi_2^\dagger) = \hat{\psi}_2 \cdot X_2 - \psi_2^\dagger \cdot X_2$, where, as above, X_2 is a vector of 1's and some or all of L_1, A_1, L_2 such that $g_2(\bar{L}_2, A_1; \psi_2) = X_2' \psi_2$.

(11)→ Substitution using equation (4).

$\bar{L}_2, A_1$ is ancillary for $\hat{\psi}_1$, so we may break the expectation of $\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger)$ into a product

of expectations as follows:

$$\begin{aligned}
E \left[\Delta_1^g(\hat{\psi}_2, \psi_2^\dagger) \right] &= E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} (I[\mathbb{P}_n IF_2 \cdot X_2 > 0] - A_2) (\mathbb{P}_n IF_2 \cdot X_2) \right] \\
&= E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} I[\mathbb{P}_n IF_2 \cdot X_2 > 0] \mathbb{P}_n IF_2 \cdot X_2 \right] \\
&\quad - E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} A_2 (\mathbb{P}_n IF_2 \cdot X_2) \right] \\
&= E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} \right] E \left[I[\mathbb{P}_n IF_2 \cdot X_2 > 0] \mathbb{P}_n IF_2 \cdot X_2 \right] \\
&\quad - E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} A_2 \right] E \left[(\mathbb{P}_n IF_2 \cdot X_2) \right] \\
&= E \left[\{S_1(A_1) - E[S_1(A_1)|L_1]\} \right] E \left[I[\mathbb{P}_n IF_2 \cdot X_2 > 0] \mathbb{P}_n IF_2 \cdot X_2 \right]
\end{aligned}$$

Note that the final equality results from the fact that $IF_2 \cdot X_2$ has mean zero.

From this, we find that the asymptotic bias of $\hat{\psi}_1$ equals

$$\begin{aligned}
& - \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} \left\{ E \left\{ I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0] \{S_1(A_1) - E[S_1(A_1)|L_1]\} \right\} E \{ I[Z_{2+} > 0] Z_{2+} \} \right. \\
& \quad \left. + E \left\{ I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) \neq 0] \right\} 0 \right\},
\end{aligned}$$

where $Z_{2+} = IF_2 \cdot X_2$ or simply

$$- \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} E \left\{ I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0] \{S_1(A_1) - E[S_1(A_1)|L_1]\} \right\} E \{ I[Z_{2+} > 0] Z_{2+} \}.$$

Thus, when optimal regimes are not unique, $I[g_2(\bar{L}_2, A_1; \psi_2^\dagger) = 0] = 1$ and the g-estimates have bounded asymptotic bias at exceptional laws equal to

$$- \left(E \left[\frac{\partial}{\partial \psi_1} U_1^*(\psi_1^\dagger, \psi_2^\dagger) \right] \right)^{-1} E \{ S_1(A_1) - E[S_1(A_1)|L_1] \} \cdot E[Z_{2+}|Z_{2+} > 0] P[Z_{2+} > 0].$$

For more than two intervals, the asymptotic bias calculation is more complex. Consider a three-interval problem where non-unique optimal rules exist at the last interval. The asymptotic bias at the second (middle) interval proceed as above, so that the asymptotic

bias in $\hat{\psi}_2$ is

$$-\left(E\left[\frac{\partial}{\partial\psi_2}U_2^*(\psi_1,\psi_2^\dagger,\psi_3^\dagger)\right]\right)^{-1}E\{S_2(A_2)-E[S_2(A_2)|L_1,A_1,L_2]\}\cdot E[Z_{3+}|Z_{3+}>0]P[Z_{3+}>0].$$

Note that in the recursive estimation setting, $U_2^*(\psi_1,\psi_2,\psi_3) = U_2^*(\psi_2,\psi_3)$. The asymptotic bias calculation at the first interval can be found by examining $\Delta_1^g(\hat{\psi}_2,\psi_2^\dagger,\hat{\psi}_3,\psi_3^\dagger) = U_1^*(\psi_1,\hat{\psi}_2,\hat{\psi}_3) - U_1^*(\psi_1,\psi_2^\dagger,\psi_3^\dagger)$. This function is more complex than that derived above because it requires plug-in estimates of both the regular estimator ψ_3 and the asymptotically biased, non-regular estimator ψ_2 .