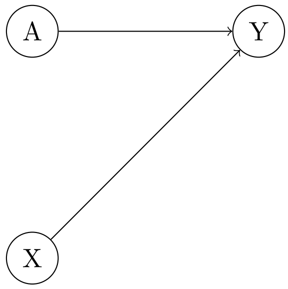
The data is modeled as:

$$\begin{aligned} (X_1, Y_1, A_1), \ldots, (X_n, Y_n, A_n) &\overset{iid}{\sim} \mathcal{O} \\ A &\perp X \\ P(A = 1 \mid X) = P(A = 1) = 1 - P(A = 0) &= p \end{aligned} \tag{1}$$

for some law $\mathcal{O}$.



The estimand is

$$\psi_0 = E(Y \mid A = 1) - E(Y \mid A = 0).$$

An estimator is obtained as the solution in ψ of

$$\sum_{i=1}^{n} U(Y_i, A_i; \psi) = 0,$$

where

$$U(Y, A; \psi) = (Y - \psi A)(A - p).$$

Consistency and asymptotic normality of this estimator follow from:

Lemma 0.1. $E(U(Y, A; \psi_0)) = 0$.

Proof.

$$\begin{aligned} E(U(Y, A; \psi_0)) &= E[(Y - \psi_0 A)(A - p)] \\ &= E[(E(Y \mid A) - \psi_0 A)(A - p)] \\ &= (E(Y \mid A = 1) - \psi_0 A)(1 - p)p + E(Y \mid A = 0)(-p)(1 - p) \\ &= p(1 - p)[E(Y \mid A = 1) - E(Y \mid A = 0) - \psi_0] = 0. \end{aligned}$$

$\square$

We consider estimators obtained as solutions in ψ to equations of the form

$$\sum_i U(Y_i, A_i; \psi) + (A_i - p)h(X_i) = 0 \tag{2}$$

for [arbitrary] functions h. It follows from Lemma 0.1 and (1) that

$$E(U(Y, A; \psi_0) + (A - p)h(X)) = 0,$$

so these estimators are also consistent and asymptotically normal. An additional benefit is that the asymptotic variance of the resulting estimator may be minimized by varying h, perhaps improving on the efficiency of the estimator obtained from $\sum_i U(Y, A; \psi) = 0$. In fact, the minimizing choice of h is determined by the estimating equation given by

$$\begin{aligned} W(\psi) &= U(\psi) - E(U(\psi) \mid S_A) \\ &= U(\psi) - E(U(\psi) \mid A, X) + E(U(\psi) \mid X). \end{aligned} \tag{3}$$

A proof is given in Lemma 0.2, after rewriting (3). The first term on the rhs of (3) is

$$\begin{aligned}E(U(\psi) \mid A, X) &= (E(Y \mid A, X) - \psi A)(-1)^{1-A} \\ &= A[E(Y \mid A=1, X) - \psi + E(Y \mid A=0, X)] - E(Y \mid A=0, X).\end{aligned}$$

For the second equality we use the identity $g(a,x) = a[g(1,x) - g(0,x)] + g(0,x)$, which holds for $a \in \{0,1\}$ and arbitrary g, x, applied to the function $g : (a,x) \mapsto (E(Y \mid a, x) - \psi a)(-1)^{1-a}$. The second term on the rhs of (3) is

$$\begin{aligned}E(U(\psi) \mid X) &= (1/2)[E(U(\psi) \mid A=1, X) + E(U(\psi) \mid A=0, X)] \\ &= (1/2)[E(Y \mid A=1, X) - \psi - E(Y \mid A=0, X)],\end{aligned}$$

using in the first equality that A and X are independent under (1). Combining the last two displays,

$$\begin{aligned}W(\psi) &:= U(\psi) - E(U(\psi) \mid A, X) + E(U(\psi) \mid X) \\ &= U(\psi) - (A - 1/2)[E(Y \mid A=1, X) + E(Y \mid A=0, X) - \psi] \\ &= U(\psi) - (A - 1/2)[2E(Y \mid X) - \psi].\end{aligned}$$

Lemma 0.2. *The asymptotic variance of the estimator obtained as the solution in* ψ *to*

$$\sum_i U(Y_i, A_i; \psi) + (A_i - 1/2)h(X_i) = 0 \tag{4}$$

is minimized over arbitrary functions h *of* X *at* $h_0(X) = -(2E(Y \mid X) - \psi_0)$.

Proof. We give the $p = P(A=1) = 1/2$ case. Under suitable regularity conditions, the asymptotic variance of the solution to the estimating equation (4) is given by the variance of the influence function,

$$-\left(E\,\frac{\partial}{\partial\psi} U(Y, A; \psi)\Big|_{\psi_0}\right)^{-1} (U(Y, A; \psi_0) + (A - 1/2)h(X)).$$

Thus we wish to show

$$\begin{aligned}\mathrm{Var}\left[\left(E\frac{\partial}{\partial\psi} U(Y, A; \psi_0)\right)^{-1} (U(Y, A; \psi_0) + (A - 1/2)h(X))\right] \geq \\ \mathrm{Var}\left[\left(E\frac{\partial}{\partial\psi} U(Y, A; \psi_0)\right)^{-1} (U(Y, A; \psi_0) + (A - 1/2)h_0(X))\right]\end{aligned}$$

or

$$\begin{aligned}E[(A - 1/2)^2 h^2(X)] + 2E[U(Y, A; \psi_0)(A - 1/2)h(X)] \geq \\ E[(A - 1/2)^2 h_0^2(X)] + 2E[U(Y, A; \psi_0)(A - 1/2)h_0(X)].\end{aligned} \tag{5}$$

Since A, X are uncorrelated, and noting that $(A - 1/2)(-1)^{1-A} = 1/2$, the lhs is

$$\begin{aligned}&E[(A - 1/2)^2 h^2(X)] + 2E[U(Y, A; \psi_0)(A - 1/2)h(X)] \\ &= \mathrm{Var}(A) E h^2(X) + 2E[(A - 1/2)E(U(Y, A; \psi_0)h(X) \mid A)] \\ &= Eh^2(X)/4 + 2E[(A - 1/2)E((Y - \psi_0 A)(-1)^{1-A} h(X) \mid A)] \\ &= Eh^2(X)/4 + E[E((Y - \psi_0 A)h(X) \mid A)] \\ &= Eh^2(X)/4 + E((Y - \psi_0/2)h(X)).\end{aligned}$$

We obtain a corresponding expression for the rhs by substituting $h(X) := h_0(X) = -(2E(Y \mid X) - \psi_0)$,

$$
\begin{aligned}
&E[(A-1/2)^2 h_0^2(X)] + 2E[U(Y,A;\psi_0)(A-1/2)h_0(X)] \\
&= Eh_0^2(X)/4 + E((Y-\psi_0/2)h_0(X)) \\
&= E[h_0(X)(h_0(X)/4 + Y - \psi_0/2)] \\
&= E[h_0(X)(-(2E(Y \mid X) - \psi_0)/4 + E(Y \mid X) - \psi_0/2)] \\
&= E[h_0(X)(E(Y \mid X)/2 - \psi_0/4)] \\
&= -Eh_0^2(X)/4 \qquad (6) \\
&= -E[E(Y \mid X)^2] + \psi_0 EY - \psi_0^2/4.
\end{aligned}
$$

Thus (5), which we wish to show, becomes

$$
Eh^2(X)/4 + E((Y-\psi_0/2)h(X)) + E[E(Y \mid X)^2] - \psi_0 EY + \psi_0^2/4 \geq 0.
$$

This inequality follows by an application of the Cauchy-Schwarz inequality,

$$
\begin{aligned}
&Eh^2(X)/4 + E((Y-\psi_0/2)h(X)) + E[E(Y \mid X)^2] - \psi_0 EY + \psi_0^2/4 \\
&= (1/4)E[(h(X)-\psi_0)^2] + E(Yh(X)) + E[E(Y \mid X)^2] - \psi_0 EY \\
&= (1/4)E[(h(X)-\psi_0)^2] + E[E(Y \mid X)^2] + E[E(Y \mid X)(h(X)-\psi_0)] \\
&\geq (1/4)E[(h(X)-\psi_0)^2] + E[E(Y \mid X)^2] - E[(h(X)-\psi_0)^2]^{1/2} E[E(Y \mid X)^2]^{1/2} \\
&= \{(1/2)E[(h(X)-\psi_0)^2]^{1/2} - E[E(Y \mid X)^2]^{1/2}\}^2 \geq 0.
\end{aligned}
$$

□

Remark. *From (6), $Eh_0^2(X)/4$ is the reduction in the asymptotic variance gained by using (4) over (2).*

From the identity $(-1)^{1-a} = (2a-1) = 2(a-1/2), a \in \{0,1\}$, we write

$$
U(\psi) = (Y - \psi A)(-1)^{1-A} = 2(A-1/2)(Y-\psi A),
$$

obtaining

$$
\begin{aligned}
W(\psi) &= U(\psi) - (A-1/2)[2E(Y \mid X) - \psi] \\
&= (A-1/2)[2(Y - E(Y \mid X)) - \psi(2A-1)] \\
&= (A-1/2)[2(Y - E(Y \mid X)) + (-1)^A \psi].
\end{aligned}
$$

[[Extension to $P(A=1 \mid X) = p \in (0,1)$]]

$$U(\psi) = (Y - \psi A)(-1)^{1-A} \quad , \quad \frac{\partial}{\partial\psi} U(\psi) = -A(-1)^{1-A} \quad , \quad \mathbb{E}\left(\frac{\partial}{\partial\psi} U(\psi)\right) = -\frac{1}{2} \quad .$$

$$n^{1/2}(\hat{\psi}_n - \psi_0) = -\left(\mathbb{E}\frac{\partial U}{\partial\psi}\Big|_{\psi_0}\right)^{-1} n^{-1/2}\sum_i U(\psi_0) = 2n^{-1/2}\sum(Y_i - \psi_0 A_i)(-1)^{1-A_i}$$

$$\tilde{Y}_i := Y_i - \mathbb{E}(Y|X_i)$$

$$\mathbb{E}(\tilde{Y}_i|A) = \alpha_0 + \beta_0 A_i + \varepsilon_i$$

$$\mathbb{E}(\tilde{Y}|A=1) = \mathbb{E}(Y|A=1) - \mathbb{E}Y = \alpha_0 + \beta_0$$

$$\mathbb{E}(\tilde{Y}|A=0) = \mathbb{E}(Y|A=0) - \mathbb{E}Y = \alpha_0$$

$$\beta_0 = \psi_0 = \mathbb{E}(Y|A=1) - \mathbb{E}(Y|A=0), \quad \alpha_0 = -\psi_0/2$$

$$0 = \sum_{i=1}^{n}\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \hat{\alpha} - A_i\hat{\beta}) = \sum_{i=1}^{n}\left\{\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \alpha_0 - A_i\beta_0) + \begin{pmatrix}-1 & -A_i\\-A_i & -A_i\end{pmatrix}\begin{pmatrix}\hat{\alpha} - \alpha_0\\\hat{\beta} - \beta_0\end{pmatrix}\right\}$$

$$n^{1/2}\begin{pmatrix}\hat{\alpha}_n - \alpha_0\\\hat{\beta}_n - \beta_0\end{pmatrix} = \left|\frac{1}{n}\sum\begin{pmatrix}1 & A_i\\A_i & A_i\end{pmatrix}\right|^{-1} n^{-1/2}\sum\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \alpha_0 - A_i\beta_0)$$

$$= \begin{pmatrix}1 & 1/2\\1/2 & 1/2\end{pmatrix}^{-1} n^{-1/2}\sum\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \alpha_0 - A_i\beta_0) = 4\begin{pmatrix}1/2 & -1/2\\-1/2 & 1\end{pmatrix} n^{-1/2}\sum\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \alpha_0 - A_i\beta_0) + o_p(1)$$

$$n^{1/2}(\hat{\beta}_n - \beta_0) = n^{-1/2}\sum(1 \; 2)\, 4\begin{pmatrix}1\\A_i\end{pmatrix}(\tilde{Y}_i - \alpha_0 - A_i\beta_0) + o_p(1) = 2n^{-1/2}\sum(2A_i - 1)(\tilde{Y}_i - \alpha_0 - A_i\beta_0) + o_p(1)$$

$$= 2n^{-1/2}\sum(-1)^{1-A_i}(\tilde{Y}_i - \alpha_0 - A_i\beta_0) + o_p(1)$$

$$n^{1/2}(\hat{\psi}_n - \hat{\beta}_n) = 2n^{-1/2}\sum(-1)^{1-A_i}(Y_i - \psi_0 A_i - \tilde{Y}_i + \alpha_0 + A_i\beta_0) + o_p(1)$$

$$= 2n^{-1/2}\sum(-1)^{1-A_i}(\mathbb{E}(Y|X_i) + \alpha_0) + o_p(1) \rightsquigarrow N(0, \sigma^2), \quad \sigma \neq 0$$