

Nonparametric doubly robust estimation of continuous treatment

This work is to verify the proof of theorems introduced in the paper (Kennedy et al, 2016).

Guide to notation

$Z = (L, A, Y)$ = observed data arising from distribution P with density $p(z) = p(y | l, a)p(a | l)p(l)$ and support $\text{supp}(Z) = \mathcal{Z} = \mathcal{L} \times \mathcal{A} \times \mathcal{Y}$

$\mathbb{P}_n = \frac{1}{n} \sum_i \delta_{z_i}$ = empirical measure so that $\mathbb{P}_n f(Z) = \frac{1}{n} \sum_i f(z_i)$

$\mathbb{P}(\hat{\mathcal{U}}) = \mathbb{P}\{\mathcal{U}(\mathbb{Z})\} = \int_{\mathcal{Z}} f(z) dP(z)$ = expectation for new Z treating f as fixed (so $\mathbb{P}(\hat{\mathcal{U}})$ is random if \hat{f} depends on sample, in which case $\mathbb{P}(\hat{\mathcal{U}}) \neq E(\hat{f})$)

$\pi(a | l) = p(a | l) = \frac{\partial}{\partial a} P(A \leq a | l)$ = conditional density of treatment A

$\hat{\pi}(a | l)$ = user-specified estimator of $\pi(a | l)$, which converges to limit $\bar{\pi}(a | l)$ that may not equal true π

$\omega(a) = p(a) = \frac{\partial}{\partial a} P(A \leq a) = E[\pi(a | L)] = \int_{\mathcal{L}} \pi(a | l) dP(l)$ = density of A

$\hat{\omega}(a) = \mathbb{P}_{\times} \{\hat{\pi}(a | l)\} = \int_{\mathcal{L}} \hat{\pi}(a | l) dP(l)$ = estimator of ω , which converges to limit $\bar{\omega}(a)$ that may not equal true ω

$\mu(l | a) = E(Y | L = l, A = a) = \int_{\mathcal{Y}} y dP(y | l, a)$ = conditional mean outcome

$\hat{\mu}(l | a)$ = user-specified estimator of $\mu(l | a)$, which converges to limit $\bar{\mu}(l, a)$ that may not equal true μ

$\hat{m}(a) = \mathbb{P}_{\times} \{\hat{\mu}(L, a)\} = \int_{\mathcal{L}} \hat{\mu}(l, a) d\mathbb{P}_{\times}(l) = \frac{1}{n} \sum_i \hat{\mu}(l_i, a)$

$\psi = \int_{\mathcal{A}} \int_{\mathcal{L}} \mu(l, a) \omega(a) dP(l) da$

Theorem 1.

Under a nonparametric model, the efficient influence function for ψ is

$\xi(Z; \pi, u) = \psi + \int_{\mathcal{A}} \{\mu(L, a) - \int_{\mathcal{L}} \mu(l, a) dP(l)\} \omega(a) da$ where

$$\xi(Z; \pi, u) = \frac{Y - \mu(L, A)}{\pi(A | L)} \int_{\mathcal{L}} \pi(A | l) dP(l) + \int_{\mathcal{L}} \mu(l, a) dP(l)$$

Importantly, the function $\xi(Z; \pi, u)$ satisfies its desired double robustness property, i.e., that $E[\xi(Z; \pi, u) | A = a] = \theta(a)$ if either $\bar{\pi} = \pi$ or $\bar{\mu} = \mu$ where $\theta(a) = E(Y^a)$

This motivates estimating the effect curve $\theta(a)$ by estimating the nuisance functions (π, μ) , and then regressing the estimated pseudo-outcome

$$\hat{\xi}(Z; \hat{\pi}, \hat{\mu}) = \frac{Y - \hat{\mu}(L, A)}{\hat{\pi}(A, L)} \int_{\mathcal{L}} \hat{\pi}(A, L) d\mathbb{P}_n(l) + \int_{\mathcal{L}} \hat{\mu}(l, a) d\mathbb{P}_n(l)$$

on treatment A using off-the-shelf nonparametric regression or machine learning methods.

Proof of Theorem 1

By definition the efficient influence function for ψ is the unique function $\phi(Z)$ that satisfies $\psi'_\epsilon(0) = E[\phi(Z)\ell'_\epsilon(Z; 0)]$, where $\psi(\epsilon)$ represents the parameter of interest as a functional on the parametric submodel and $\ell(w | \bar{w}; \epsilon) = \log p(w | \bar{w}; \epsilon)$ for any partition $(W, \bar{W}) \subseteq Z$. Therefore

$$\ell'_\epsilon(z; \epsilon) = \ell'_\epsilon(y | l, a; \epsilon) + \ell'_\epsilon(a | l; \epsilon) + \ell'_\epsilon(l; \epsilon)$$

The authors give two important properties of such score functions of such score functions $\ell'_\epsilon(w | \bar{w}; \epsilon)$ that will be used throughout this proof. First note that since $\ell_\epsilon(w | \bar{w}; \epsilon)$ is a log transformation of $p(w | \bar{w}; \epsilon)$, it follows that $\ell'_\epsilon(w | \bar{w}; \epsilon) = p'(w | \bar{w}; \epsilon)/p(w | \bar{w}; \epsilon)$. Similarly, as with any score function, note that $E[\ell'_\epsilon(W | \bar{W}; \epsilon) | \bar{W}] = 0$ since

$$\int_{\mathcal{W}} \ell'_\epsilon(w | \bar{w}; 0) dP(w | \bar{w}) = \int_{\mathcal{W}} dP'(w | \bar{w}) = \frac{\partial}{\partial \epsilon} \int_{\mathcal{W}} P(w | \bar{w}) = 0$$

The goal in this proof is to show that $\psi'_\epsilon(0) = E[\phi(Z)]\ell'_\epsilon(Z; 0)$ for the proposed influence function $\phi(Z) = \xi(Z; \pi, u) - \psi + \int_{\mathcal{A}} \{\mu(L, a) - \int_{\mathcal{L}} \{\mu(l, a) dP(l)\} \omega(a) da$. First we will give an expression for $\psi'_\epsilon(0)$. By definition $\psi(\epsilon) = \int_{\mathcal{A}} \theta(a; \epsilon) \omega(a; \epsilon) da$ because

$$\begin{aligned} \psi &= \int_{\mathcal{A}} \int_{\mathcal{L}} \mu(l, a) \omega(a) dP(l) da = \int_{\mathcal{A}} \int_{\mathcal{L}} \mu(l, a) dP(l) dP(a) \\ &= E[E[\mu(l, a)]] = E[E(Y | L = l, A = a)] = E[E[Y^a]] = E[\theta(a)]. \end{aligned}$$

Hence, $\psi'_\epsilon(0) = \int_{\mathcal{A}} \{\theta'(a; 0) \omega(a) + \theta(a) \omega'(a; 0)\} da = E[\theta'_\epsilon(A; 0) + \theta(A) \ell'_\epsilon(A; 0)]$ because

$$\omega(a) = p(a) \text{ and } \ell'_\epsilon(A; 0) = \frac{P'_\epsilon(A; 0)}{P(A; 0)} \iff \omega'_\epsilon(A; 0) = \ell'_\epsilon(A; 0) P(A; 0)$$

Also since $\theta(a; \epsilon) = \int_{\mathcal{L}} \int_{\mathcal{Y}} y p(y | l, a; \epsilon) p(l; \epsilon) d\eta(y) dv(l)$, we have

$$\begin{aligned} \theta'(a; 0) &= \int_{\mathcal{L}} \int_{\mathcal{Y}} y \{p'_\epsilon(y | l, a; 0) p(l) + p(y | l, a) p'(l; 0)\} d\eta(y) dv(l) \\ &= \theta'(a; 0) = \int_{\mathcal{L}} \int_{\mathcal{Y}} y \{\ell'_\epsilon(y | l, a; 0) p(y | l, a) p(l) + p(y | l, a) \ell'_\epsilon(l; 0) p(l)\} d\eta(y) dv(l) \\ &= E[E\{Y \ell'_\epsilon(Y | L, A; 0) | L, A = a\}] + E[\mu(L, a) \ell'_\epsilon(L; 0)] \end{aligned}$$

Therefore

$$\psi'_\epsilon(0) = \int_{\mathcal{A}} \left(E[E\{Y \ell'_\epsilon(Y | L, A; 0) | L, A = a\}] + E[\mu(L, a) \ell'_\epsilon(L; 0)] + \theta(a) \ell'_\epsilon(a; 0) \right) \omega(a) da$$

Now we will consider the covariance

$$E[\phi(Z) \ell'_\epsilon(Z; 0)] = E[\phi(Z) \{\ell'_\epsilon(Y | L, A; 0) + \ell'_\epsilon(A, L; 0)\}] \text{ because}$$

$$\ell'_\epsilon(z; \epsilon) = \ell'_\epsilon(y | l, a; \epsilon) + \ell'_\epsilon(a | l; \epsilon) + \ell'_\epsilon(l; \epsilon) = \ell'_\epsilon(y | l, a; 0) + \ell'_\epsilon(a, l; 0)$$

which we need to show equals the earlier expression for $\psi'_\epsilon(0)$.

The proposed efficient influence function given in **Theorem 1** is

$$\xi(Z; \pi, u) = \frac{Y - \mu(L, A)}{\pi(A | L)} \int_{\mathcal{L}} \pi(A | l) dP(l) + \int_{\mathcal{L}} \mu(l, a) dP(l)$$

where we define $m(a) = \int_{\mathcal{L}} \mu(l, a) dP(l)$ as the marginalized version of the regression function μ , so that $m(a) = \theta(a)$ if μ is the true regression function.

$$m(a) = \int_{\mathcal{L}} \mu(l, a) dP(l) = \int_{\mathcal{L}} \int_{\mathcal{Y}} y dP(y | l, a) dP(l) = \int_{\mathcal{L}} \int_{\mathcal{Y}} y p(y | l, a) p(l) d\eta(y) dv(l) = \theta(a)$$

$$\begin{aligned}
& \text{Thus, } E[\phi(Z)\ell'_\epsilon(Y \mid L, A; 0)] \\
&= E\left(\left[\frac{Y-\mu(L,A)}{\pi(A|L)/\omega(A)} + \int_{\mathcal{A}}\{\mu(L,a)-\theta(a)\}\omega(a) \, da + \theta(A) - \psi\right]\ell'_\epsilon(Y \mid L, A; 0)\right) \\
&= E\left[\frac{Y\ell'_\epsilon(Y|L,A;0)}{\pi(A|L)/\omega(A)}\right] \\
&= E\left[\frac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/\omega(A)}\right] \\
&= \int_{\mathcal{A}} E[E\{Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A = a\}]\omega(a) \, da
\end{aligned}$$

where the first equality follows since

$$\begin{aligned}
\phi(Z) &= \frac{Y-\mu(L,A)}{\pi(A|L)} \int_{\mathcal{L}} \pi(A \mid l) dP(l) + \int_{\mathcal{L}} \mu(l, a) dP(l) - \psi + \int_{\mathcal{A}} \{\mu(L, a) - \int_{\mathcal{L}} \mu(l, a) dP(l)\} \omega(a) da \\
&= \frac{Y-\mu(L,A)}{\pi(A|L)/\omega(A)} + \theta(A) - \psi + \int_{\mathcal{A}} \{\mu(L, a) - \theta(A)\} \omega(a) da
\end{aligned}$$

the second since $E[\ell'_\epsilon(Y \mid L, A; 0) \mid L, A] = 0$, the third by iterated expectation conditioning on L and A, the last by iterated expectation conditioning on L gives:

$$\begin{aligned}
& E\left(E\left[\frac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/\omega(A)} \mid L\right]\right) \\
&= E\left(E\left[\frac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/E[\pi(A|L)]} \mid L\right]\right) \\
&= E\left(E\left[E[Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A]\right]\right) \\
&= \int_{\mathcal{A}} E[E\{Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A = a\}]\omega(a) \, da
\end{aligned}$$

Now, $E[\phi(Z)\ell'_\epsilon(A, L; 0)]$

$$\begin{aligned}
&= E\left[\left\{\frac{Y-\mu(L,A)}{\pi(A|L)/\omega(A)}\right\}\ell'_\epsilon(A, L; 0) + \{\theta(A) - \psi\}\{\ell'_\epsilon(L \mid A; 0) + \ell'_\epsilon(A; 0)\}\right. \\
&\quad \left.+ \int_{\mathcal{A}} \{\mu(L, a) - \theta(a)\}\omega(a) da \{\ell'_\epsilon(A \mid L; 0) + \ell'_\epsilon(L; 0)\}\right] \\
&= E\left[\theta(A)\epsilon(A; 0) + \int_{\mathcal{A}} \mu(L, a)\ell'_\epsilon(L; 0)\omega(a) da\right]
\end{aligned}$$

since by definition $\ell'_\epsilon(A, L; 0) = \ell'_\epsilon(L \mid A; 0) + \ell'_\epsilon(A; 0) = \ell'_\epsilon(A \mid L; 0) + \ell'_\epsilon(L; 0)$, and the equality used iterated expectation conditioning on L and A for the first term in the first line:

$$E[\ell'_\epsilon(A, L; 0) \mid A, L] = \int_{\mathcal{L}} \ell'_\epsilon(A, l; 0) dP(A, l) = \int_{\mathcal{L}} dP'_\epsilon(A, l) = \frac{\partial}{\partial \epsilon} \int_{\mathcal{L}} dP(A, L) = 0$$

iterated expectation conditioning on A for the second term in the first line:

$$E[\ell'_\epsilon(L \mid A; 0) \mid A] = 0 \text{ and } E[\ell'_\epsilon(A; 0) \mid A] = 0$$

and iterated expectation conditioning on L for the the second line.

Adding the expression $E[\phi(Z)\ell'_\epsilon(Y \mid L, A; 0)]$ and $E[\phi(Z)\ell'_\epsilon(A, L; 0)]$ gives:

$$\int_{\mathcal{A}} \left(E[E\{Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A = a\} + \mu(L, a)\ell'_\epsilon(L; 0)] + \theta(A)\epsilon(A; 0)\right)\omega(a) \, da = \psi'_\epsilon(0)$$

$\therefore \phi$ is the efficient influence function.

Double robutness of efficient influence function & mapping

Here the goal is to show that $E[\phi(Z; \bar{\pi}, \bar{\mu}, \psi)] = 0$ if either $\bar{\pi} = \pi$ or $\bar{\mu} = \mu$, where $\phi(Z; \bar{\pi}, \bar{\mu}, \psi)$ is the influence function defined as

$$\xi(Z; \bar{\pi}, \bar{\mu}) - \psi + \int_{\mathcal{A}} \{\bar{\mu}(L, a) - \int_{\mathcal{L}} \bar{\mu}(l, a) dP(l)\} \int_{\mathcal{L}} \bar{\pi}(a, l) dP(l) da,$$

where

$$\xi(Z; \bar{\pi}, \bar{\mu}) = \frac{Y - \bar{\mu}(L, A)}{\bar{\pi}(A|L)} \int_{\mathcal{L}} \bar{\pi}(A | l) dP(l) + \int_{\mathcal{L}} \bar{\mu}(l | A) dP(l)$$

First note that, letting $\bar{\omega}(a) = E[\bar{\pi}(a | L)]$ and $\bar{m}(a) = E[\bar{\mu}(L, a)]$, we have

$$\begin{aligned} & E[\xi(Z; \bar{\pi}, \bar{\mu}) | A = a] \\ &= E\left[\frac{Y - \bar{\mu}(L, A)}{\bar{\pi}(A|L)/\bar{\omega}(A)} + \bar{m}(A) | A = a\right] \\ &= \int_{\mathcal{L}} \frac{\mu(l, a) - \bar{\mu}(l, a)}{\bar{\pi}(a | l)/\bar{\omega}(a)} dP(l | a) + \bar{m}(a) \\ &= \int_{\mathcal{L}} \{\mu(l, a) - \bar{\mu}(l, a)\} \frac{\pi(a | l)/\omega(a)}{\bar{\pi}(a | l)/\bar{\omega}(a)} dP(l) + \bar{m}(a) \\ &= \theta(a) + \int_{\mathcal{L}} \mu(l, a) - \bar{\mu}(l, a) \left\{ \frac{\pi(a | l)/\omega(a)}{\bar{\pi}(a | l)/\bar{\omega}(a)} - 1 \right\} dP(l) \end{aligned}$$

where the first equality follows by definition, the second equality follows because:

$$\int_{\mathcal{L}} \mu(l, a) dP(l | a) = E[u(l, a)] = E[Y^a]$$

the third follows since $p(l | a) = p(a | l)p(l)/p(a)$ and the fourth by rearranging:

$$\theta(a) - \int_{\mathcal{L}} \{\mu(l, a) - \bar{\mu}(l, a)\} dP(l) = \int_{\mathcal{L}} \mu(l, a) dP(l) - \int_{\mathcal{L}} \{\mu(l, a) - \bar{\mu}(l, a)\} dP(l) = \bar{m}(a)$$

The last line shows that $E[\xi(Z; \bar{\pi}, \bar{\mu}) | A = a] = \theta(a)$ as long as either $\bar{\pi} = \pi$ or $\bar{\mu} = \mu$, since in either case the remainder is zero.

Therefore if $\bar{\pi} = \pi$ or $\bar{\mu} = \mu$, we have:

$$\int_{\mathcal{A}} E[\xi(Z; \bar{\pi}, \bar{\mu}) | A = a] \omega(a) da - \psi = \int_{\mathcal{A}} \theta(a) \omega(a) da - \psi$$

so that

$$\begin{aligned} & E[\phi(Z; \bar{\pi}, \bar{\mu}, \psi)] = E\left[\int_{\mathcal{A}} \{\bar{\mu}(L, a) - \bar{m}(a)\} \bar{\omega}(a) da\right] \text{ because} \\ & E[\phi(Z; \bar{\pi}, \bar{\mu}, \psi)] \\ &= E\left[\xi(Z; \bar{\pi}, \bar{\mu}) - \psi + \int_{\mathcal{A}} \{\bar{\mu}(L, a) - \int_{\mathcal{L}} \bar{\mu}(l, a) dP(l)\} \int_{\mathcal{L}} \bar{\pi}(a, l) dP(l) da\right] \\ &= E\left[E[\xi(Z; \bar{\pi}, \bar{\mu}) | A] - \psi + \int_{\mathcal{A}} \{\bar{\mu}(L, a) - \int_{\mathcal{L}} \bar{\mu}(l, a) dP(l)\} \int_{\mathcal{L}} \bar{\pi}(a, l) dP(l) da\right] \\ &= E\left[0 + \int_{\mathcal{A}} \{\bar{\mu}(L, a) - \bar{m}(a)\} \bar{\omega}(a) da\right] \end{aligned}$$

And

$$E\left[\int_{\mathcal{A}} \{\bar{\mu}(L, a) - \bar{m}(a)\} \bar{\omega}(a) da\right] = \int_{\mathcal{A}} \{\bar{m}(a) - \bar{m}(a)\} \bar{\omega}(a) da = 0 \text{ by definition:}$$

$$\begin{aligned} & E\left[\int_{\mathcal{A}} \{\bar{\mu}(L, a) - \bar{m}(a)\} \bar{\omega}(a) da\right] = \int_{\mathcal{L}} \int_{\mathcal{A}} \{\bar{\mu}(L, a) - \bar{m}(a)\} \bar{\omega}(a) dadl \\ &= \int_{\mathcal{A}} \int_{\mathcal{L}} \{\bar{\mu}(L, a) - \bar{m}(a)\} dl \bar{\omega}(a) da \text{ (Fubini's theorem)} = \int_{\mathcal{A}} E[\bar{\mu}(L, a) - \bar{m}(a)] \bar{\omega}(a) da \\ &= \int_{\mathcal{A}} \{\bar{m}(a) - \bar{m}(a)\} \bar{\omega}(a) da = 0 \\ &\therefore E[\phi(Z; \bar{\pi}, \bar{\mu}, \psi)] = 0 \text{ if either } \bar{\pi} = \pi \text{ or } \bar{\mu} = \mu \end{aligned}$$

Reference

Kennedy, E. H., Ma, Z., McHugh, M. D., & Small, D. S. (2017). Nonparametric methods for doubly robust estimation of continuous treatment effects. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 79(4), 1229–1245. <https://doi.org/10.1111/rssb.12212>