Nonparametric dobuly robust estimation of continuous treatment

This work is to verify the proof of theorems introduced in the paper (Kennedy et al, 2017).

Guide to notation

$Z = (L, A, Y)$ = observed data arising from distribution P with density $p(z) = p(y \mid l, a)p(a \mid l)p(l)$ and support 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 $\mathbb{P}_n f(Z) = \frac{1}{n}\sum_i f(z_i)$

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

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

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

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

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

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

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

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

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

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

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

Importantly, the function $\xi(Z; \pi, u)$ satisfies its desired double robustness property,
i.e., that $E[\xi(Z; \pi, u) \mid A = a] = \theta(a)$ if either $\overline{\pi} = \pi$ or $\overline{\mu} = \mu$ 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 psuedo-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 $\ell(w \mid \overline{w};\epsilon) = log\ p(w \mid \overline{w};\epsilon)$ for any partition $(W, \overline{W}) \subseteq Z$. Therefore

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

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

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

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

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

$$= E\big[E[\mu(l,a)]\big] = E[E(Y \mid L = l, A = a)] = E\big[E[Y^a]\big] = E[\theta(a)].$$

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

$$\theta'(a;0) = \int_{\mathcal{L}}\int_{\mathcal{Y}} y\ \{p'_\epsilon(y \mid l, a;0)p(l) + p(y \mid l, a)p'(l;0)\}\ d\eta(y)\ dv(l)$$

$$= \theta'(a;0) = \int_{\mathcal{L}}\int_{\mathcal{Y}} y\ \{\ell'_\epsilon(y \mid l, a;0)p(y \mid l, a)p(l) + p(y \mid l, a)\ell'_\epsilon(l;0)p(l)\}\ d\eta(y)\ dv(l)$$

$$= E\big[E\{Y\ell'_\epsilon(Y \mid L, A;0) \mid L, A = a\}\big] + E\big[\mu(L,a)\ell'_\epsilon(L;0)\big]$$

Therefore

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

Now we will consider the covariance

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

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

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

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

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

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.

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

Thus, $E\big[\phi(Z)\ell'_\epsilon(Y \mid L, A; 0)\big]$

$$= E\Big(\Big[\tfrac{Y-\mu(L,A)}{\pi(A|L)/\omega(A)} + \int_{\mathcal{A}}\{\mu(L, a - \theta(a)\}\omega(a)\ da + \theta(A) - \psi\Big]\ell'_\epsilon(Y \mid L, A; 0)\Big)$$

$$= E\Big[\tfrac{Y\ell'_\epsilon(Y|L,A;0)}{\pi(A|L)/\omega(A)}\Big]$$

$$= E\Big[\tfrac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/\omega(A)}\Big]$$

$$= \int_{\mathcal{A}} E\big[E\{Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A = a\}\big]\omega(a)\ da$$

where the first equality follows since

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

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

$$E\Big(E\Big[\tfrac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/\omega(A)} \mid L\Big]\Big)$$

$$= E\Big(E\Big[\tfrac{E[Y\ell'_\epsilon(Y|L,A;0)|L,A]}{\pi(A|L)/E[\pi(A|L)]} \mid L\Big]\Big)$$

$$= E\Big(E\Big[E[Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A]\Big]\Big)$$

$$= \int_{\mathcal{A}} E\big[E\{Y\ell'_\epsilon(Y \mid L, A; 0) \mid L, A = a\}\big]\omega(a)\ da$$

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

$$= E\Big[\Big\{\tfrac{Y-\mu(L,A)}{\pi(A|L)/\omega(A)}\Big\}\ell'_\epsilon(A, L; 0) + \{\theta(A) - \psi\}\big\{\ell'_\epsilon(L \mid A; 0) + \ell'_\epsilon(A; 0)\big\}$$
$$+ \int_{\mathcal{A}}\big\{\mu(L,a) - \theta(a)\big\}\omega(a)da\big\{\ell'_\epsilon(A \mid L; 0) + \ell'_\epsilon(L; 0)\big\}\Big]$$

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

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:

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

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

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

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

Double robutness of efficient influence function & mapping

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

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

where

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

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

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

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

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

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

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

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

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

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

so that

$$E[\phi(Z;\overline{\pi},\overline{\mu},\psi)]=E\big[\textstyle\int_{\mathcal{A}}\big\{\overline{\mu}(L,a)-\overline{m}(a)\big\}\overline{\omega}(a)da\big]\ \text{because}$$

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

And

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

$$\begin{aligned}
&E\big[\textstyle\int_{\mathcal{A}}\big\{\overline{\mu}(L,a)-\overline{m}(a)\big\}\overline{\omega}(a)da\big]=\int_{\mathcal{L}}\int_{\mathcal{A}}\big\{\overline{\mu}(L,a)-\overline{m}(a)\big\}\overline{\omega}(a)\ dadl\\
&=\ \textstyle\int_{\mathcal{A}}\int_{\mathcal{L}}\big\{\overline{\mu}(L,a)-\overline{m}(a)\big\}dl\ \overline{\omega}(a)da(Fubini's\ theorem)=\int_{\mathcal{A}}E\big[\overline{\mu}(L,a)-\overline{m}(a)\big]\overline{\omega}(a)da\\
&=\ \textstyle\int_{\mathcal{A}}\big\{\overline{m}(a)-\overline{m}(a)\big\}\overline{\omega}(a)da=0
\end{aligned}$$

$\therefore E[\phi(Z;\overline{\pi},\overline{\mu},\psi)]=0$ if either $\overline{\pi}=\pi$ or $\overline{\mu}=\mu$

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