Statistica Sinica

A stable and more efficient doubly robust estimator

Min Zhang[1*] and Baqun Zhang[2**]

[1] *University of Michigan, Ann Arbor, US,*

[2] *Shanghai University of Finance and Economics, Shanghai, P.R.China*

* *email:* mzhangst@umich.edu

** *email:* zhang.baqun@mail.shufe.edu.cn

Appendix A: Double-robustness and Consistency:

For simplicity, we denote $\pi(X_i;\widehat{\theta})$ by $\widehat{\pi}_i$ and $K\big(\frac{\widehat{\pi}_j-\widehat{\pi}_i}{h_n}\big) \equiv K\big(\frac{\pi(X_j;\widehat{\theta})-\pi(X_i;\widehat{\theta})}{h_n}\big)$ by $K_{ij}(\widehat{\theta})$. We assume $\widehat{\beta} \xrightarrow{p} \beta^*$, and $\widehat{\theta} \xrightarrow{p} \theta^*$, which is true under some regularity conditions by standard M-estimation theory (Tsiatis, 2006).

Part 1: Suppose the model for outcome is correct, i.e., $E(Y|R=1,X) = m(X;\beta)$. Then $\beta^* = \beta_0$, where β_0 is the truth, and $n^{-1}\sum_{i=1}^n m(X_i;\widehat{\beta}) \xrightarrow{p} n^{-1}\sum_{i=1}^n E(Y_i|X_i, R=1) = n^{-1}\sum_{i=1}^n E(Y_i|X_i) = E(Y)$, where the first equality is due to $Y_i \perp\!\!\!\perp R_i | X_i$. The first term of

(2.3) in the main manuscript can be decomposed as follows

$$= \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\sum_{j=1}^{n} R_j\{Y_j - m(X_j;\widehat{\beta})\}K_{ij}(\widehat{\theta})}{\sum_{j=1}^{n} R_j K_{ij}(\widehat{\theta})} - \frac{\sum_{j=1}^{n} R_j\{Y_j - m(X_j;\beta_0)\}K_{ij}(\widehat{\theta})}{\sum_{j=1}^{n} R_j K_{ij}(\widehat{\theta})}\Big] \tag{A.1}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\sum_{j=1}^{n} R_j\{Y_j - m(X_j;\beta_0)\}K_{ij}(\widehat{\theta})}{\sum_{j=1}^{n} R_j K_{ij}(\widehat{\theta})} - \frac{\sum_{j=1}^{n} R_j\{Y_j - m(X_j;\beta_0)\}K_{ij}(\theta^*)}{\sum_{j=1}^{n} R_j K_{ij}(\theta^*)}\Big] \tag{A.2}$$

$$+ \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\sum_{j=1}^{n} R_j\{Y_j - m(X_j;\beta_0)\}K_{ij}(\theta^*)}{\sum_{j=1}^{n} R_j K_{ij}(\theta^*)}\Big]. \tag{A.3}$$

We repeatedly use two standard results from kernel density and kernel nonparametric regression estimator. That is, under standard conditions usually assumed for $K(u)$, including $\int K(u)du = 1$, $\int uK(u)du = 0$, $h_n \to 0$, and $nh_n \to \infty$, we have $n^{-1}\sum_{i=1}^{n} \frac{1}{h_n}K(\frac{x-X_i}{h_n}) \xrightarrow{p} f_X(x)$ and $n^{-1}\sum_{i=1}^{n}\{Y_i \frac{1}{h_n}K(\frac{x-X_i}{h_n})\} \xrightarrow{p} E(Y|x)f_X(x)$, where $f_X(x)$ is the density of X. As a result, $\frac{n^{-1}\sum_{i=1}^{n} Y_i K(\frac{x-X_i}{h_n})}{n^{-1}\sum_{i=1}^{n} K(\frac{x-X_i}{h_n})} \xrightarrow{p} E(Y|x)$. Applying these results, we have $(A.3) \xrightarrow{p} E[E\{Y - m(X;\beta_0)|R = 1, \theta^{*T}X\}]$. The inner expectation can be written equivalently as

$$\begin{aligned}
& E\{Y - m(X;\beta_0)|R = 1, \theta^{*T}X\} \\
= \;& E[E\{Y - m(X;\beta_0)|R = 1, \theta^{*T}X, X\}|R = 1, \theta^{*T}X] \\
= \;& E[E\{Y - m(X;\beta_0)|R = 1, X\}|R = 1, \theta^{*T}X] = 0
\end{aligned}$$

due to $E(Y|R = 1, X) = m(X;\beta_0)$. Therefore, (A.3)$\xrightarrow{p}$ 0. It can be shown that $(A.1) = (\beta_0 - \widehat{\beta})n^{-1}\sum_{i=1}^{n}\Big[\frac{\sum_{j=1}^{n} R_j K_{ij}(\widehat{\theta})\{\frac{d}{d\beta}|_{\beta=\beta_0} m(X_j;\beta)\}}{\sum_{j=1}^{n} R_j K_{ij}(\widehat{\theta})}\Big] = o_p(1)$. It can also be checked that (A.2) $\xrightarrow{p}$ 0. Combining results, we obtain that when the model for outcome is correctly specified, $\mu \xrightarrow{p} \mu = E(Y)$.

Part 2: Suppose the model for the propensity score is correct, i.e., $P(R = 1|X) = \pi(X;\theta)$. Then $\theta^* = \theta_0$, where θ_0 is the truth, and $Y \perp\!\!\!\perp R|\theta_0^T X$ or equivalently $Y \perp\!\!\!\perp R|\pi(X;\theta_0)$ (Rosenbaum and Rubin, 1983). Rearranging terms we can write $\widehat{\mu}$ equivalently as

$$n^{-1}\sum_{i=1}^{n}\Big\{\frac{\sum_{j=1}^{n}R_jY_jK_{ij}(\widehat{\theta})}{\sum_{j=1}^{n}R_jK_{ij}(\widehat{\theta})}\Big\}+n^{-1}\sum_{i=1}^{n}\Big\{m(X_i;\widehat{\beta})-\frac{\sum_{j=1}^{n}R_jm(X_j;\widehat{\beta})K_{ij}(\widehat{\theta})}{\sum_{j=1}^{n}R_jK_{ij}(\widehat{\theta})}\Big\}. \tag{A.4}$$

The first term in (A.4)

$$\begin{aligned} &= n^{-1}\sum_{i=1}^{n}\Big\{\frac{\sum_{j=1}^{n}R_jY_jK_{ij}(\theta_0)}{\sum_{j=1}^{n}R_jK_{ij}(\theta_0)}\Big\} && \text{(A.5)}\\ &+ n^{-1}\sum_{i=1}^{n}\Big\{\frac{\sum_{j=1}^{n}R_jY_jK_{ij}(\widehat{\theta})}{\sum_{j=1}^{n}R_jK_{ij}(\widehat{\theta})}\Big\}-n^{-1}\sum_{i=1}^{n}\Big\{\frac{\sum_{j=1}^{n}R_jY_jK_{ij}(\theta_0)}{\sum_{j=1}^{n}R_jK_{ij}(\theta_0)}\Big\}. && \text{(A.6)}\end{aligned}$$

Applying results on kernel estimator and law of large numbers, we have (A.5) $\xrightarrow{p} E\{E(Y|R=1,\theta_0^TX)\}=E\{E(Y|\theta_0^TX)\}=E(Y)$, because $Y\perp\!\!\!\perp R|\theta_0^TX$. Because $\widehat{\theta}\xrightarrow{p}\theta_0$ and $\pi(x;\widehat{\theta})\xrightarrow{p}\pi(x;\theta_0)$, it can be shown that (A.6) $\xrightarrow{p} 0$. The second term in (A.4)

$$\begin{aligned} &= n^{-1}\sum_{i=1}^{n}\Big\{m(X_i;\beta^*)-\frac{\sum_{j=1}^{n}R_jm(X_j;\beta^*)K_{ij}(\theta_0)}{\sum_{j=1}^{n}R_jK_{ij}(\theta_0)}\Big\}+o_p(1)\\ &\xrightarrow{p} E\{m(X;\beta^*)\}-E\big[\frac{E\{Rm(X;\beta^*)|\theta_0^TX\}}{E(R|\theta_0^TX)}\big]\end{aligned}$$

due to law of large numbers and results on kernel estimators. By Rosenbaum and Rubin (1983), $R\perp\!\!\!\perp X|\theta_0X$, and therefore, $E\{Rm(X;\beta^*)|\theta_0^TX\}=E(R|\theta_0^TX)E\{m(X;\beta^*)|\theta_0^TX\}$. Substituting this result back, the second term in (A.4) $\xrightarrow{p} E\{m(X;\beta^*)\}-E\big[E\{m(X;\beta^*)|\theta_0^TX\}\big]=0$.

Combining results, we obtain that when the model for the propensity score is correctly specified, regardless of whether the model for outcome is correct or not, $\widehat{\mu}\xrightarrow{p}\mu=E(Y)$.

Appendix B: Asymptotic Normality:

We first define the following quantities:

$$
\begin{aligned}
& g_1(r,y,x;\tilde{x},\beta,\theta) = r\{y-m(x;\beta)\}K(\frac{\pi(x;\theta)-\pi(\tilde{x};\theta)}{h_n}), \\
& g_2(r,x;\tilde{x},\theta) = rK(\frac{\pi(x;\theta)-\pi(\tilde{x};\theta)}{h_n}), \\
& g_3(\tilde{x};\beta,\theta) = \frac{P_n g_1(R,Y,X;\tilde{x},\theta)}{P_n g_2(R,X;\tilde{x},\beta,\theta)}, \quad g_4(r,y,x;\beta,\theta) = P\{\frac{g_1(r,y,x;\tilde{x},\beta,\theta)}{P_n g_2(R,X;\tilde{x},\theta)}|_{\tilde{x}=X}\} \\
& g_5(r,x;\beta,\theta) = P\{\frac{g_2(r,x;\tilde{x},\theta)Pg_1(R,Y,X;\tilde{x},\beta,\theta)}{P_n g_2(R,X;\tilde{x},\theta)Pg_2(R,X;\tilde{x},\theta)}|_{\tilde{x}=X}\}.
\end{aligned}
$$

With these notations, we have $\widehat{\mu} = P_n g_3(X;\widehat{\beta},\widehat{\theta}) + P_n m(X;\widehat{\beta})$ and can be decomposed as

$$
\begin{aligned}
\widehat{\mu} &= (P_n - P)g_3(X;\widehat{\beta},\widehat{\theta}) + (P_n - P)g_4(R,Y,X;\widehat{\beta},\widehat{\theta}) - (P_n - P)g_5(R,X;\widehat{\beta},\widehat{\theta}) \\
&+ P\Big\{\frac{Pg_1(R,Y,X;\tilde{x},\widehat{\beta},\widehat{\theta})}{Pg_2(R,X;\tilde{x},\widehat{\theta})}|_{\tilde{x}=X}\Big\} + P_n m(X;\widehat{\beta}).
\end{aligned}
$$

It can be shown that

$$
\begin{aligned}
G_n g_3(X;\widehat{\beta},\widehat{\theta}) &= G_n \frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}} + o_p(1), \\
G_n g_4(R,Y,X;\widehat{\beta},\widehat{\theta}) &= G_n \frac{R\{Y-m(X;\beta^*)\}}{E\{R|\pi(X;\theta^*)\}} + o_p(1), \qquad \text{(B.1)} \\
G_n g_5(R,X;\widehat{\beta},\widehat{\theta}) &= G_n \frac{RE[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E^2\{R|\pi(X;\theta^*)\}} + o_p(1). \qquad \text{(B.2)}
\end{aligned}
$$

Specifically, in addition to the two results on kernel density and kernel nonparametric regression estimator presented before, for proving (B.1) we additionally need the following result:

$$
\begin{aligned}
& P\Big[\frac{K(\frac{z-\tilde{z}}{h_n})}{P_n\{RK(\frac{Z-\tilde{z}}{h_n})\}}|_{\tilde{z}=Z}\Big] = \int \frac{\frac{1}{h_n}K(\frac{t-z}{h_n})f_Z(t)}{E(R|Z=t)f_Z(t)}dt + o_p(1) \\
&= \int \frac{K(u)}{E(R|Z=z+uh_n)}du + o_p(1) \\
&= \int \big[\frac{K(u)}{E(R|Z=z)} + uh_n K(u)\frac{d\{\frac{1}{E(R|Z=z)}\}}{dz} + o(h_n)\big]du + o_p(1) \\
&= \frac{1}{E(R|Z=z)} + o_p(1),
\end{aligned}
$$

due to Taylor expansion and standard conditions assumed for $K(u)$, including $\int K(u)du = 1$, $\int uK(u)du = 0$, $h_n \to 0$, and $nh_n \to \infty$. Similar techniques are used for proving (B.2).

Moreover,

$$
\begin{aligned}
&n^{\frac{1}{2}}(B.1) = n^{\frac{1}{2}} E\Big\{\frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}}\Big\} + n^{\frac{1}{2}}(P_n - P)m(X;\beta^*) \\
&+\frac{d}{d\theta}|_{\theta=\theta^*} E\Big[\frac{R\{Y-m(X;\beta^*)\}}{E\{R|\pi(X;\theta)\}}\Big] n^{\frac{1}{2}}(\widehat{\theta}-\theta^*) \\
&+\frac{d}{d\beta}|_{\beta=\beta^*} E\Big[\frac{R\{Y-m(X;\beta)\}}{E\{R|\pi(X;\theta^*)\}} + m(X;\beta)\Big] n^{\frac{1}{2}}(\widehat{\beta}-\beta^*),
\end{aligned}
$$

where we have used the result that $E\Big\{\frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}}\Big\} = E\Big[\frac{R\{Y-m(X;\beta^*)\}}{E\{R|\pi(X;\theta^*)\}}\Big]$.

Combining results, we obtain

$$
\begin{aligned}
&n^{\frac{1}{2}}(\widehat{\mu}-\mu) = n^{\frac{1}{2}}(P_n - P)\Big\{\frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}} \\
&+\frac{R\{Y-m(X;\beta^*)\}}{E\{R|\pi(X;\theta^*)\}} - \frac{RE[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E^2\{R|\pi(X;\theta^*)\}} + m(X;\beta^*) - \mu\Big\} \\
&+\frac{d}{d\theta}|_{\theta=\theta^*} E\Big[\frac{R\{Y-m(X;\beta^*)\}}{E\{R|\pi(X;\theta)\}}\Big] n^{\frac{1}{2}}(\widehat{\theta}-\theta^*) \\
&+\frac{d}{d\beta}|_{\beta=\beta^*} E\Big[\frac{R\{Y-m(X;\beta)\}}{E\{R|\pi(X;\theta^*)\}} + m(X;\beta)\Big] n^{\frac{1}{2}}(\widehat{\beta}-\beta^*) \\
&+n^{\frac{1}{2}}\Big[E\Big\{\frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}} + m(X;\beta^*)\Big\} - \mu\Big] + o_p(1).
\end{aligned}
$$

It was checked previously in Appendix A that, when either working model for Y or R is correctly specified, then the asymptotic bias $E\Big\{\frac{E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta^*)]}{E\{R|\pi(X;\theta^*)\}} + m(X;\beta^*)\Big\} - \mu$ equals zero. Under suitable regularity conditions, according to M-estimation theory it is standard results that $n^{\frac{1}{2}}(\widehat{\beta}-\beta^*)$ and $n^{\frac{1}{2}}(\widehat{\theta}-\theta^*)$ are asymptotic normal with mean zero. Therefore, the consistency and asymptotic normality of $\widehat{\mu}$ hold.

Specifically, when $\pi(X;\theta)$ is the correct model for $R|X$, then $\widehat{\theta} \xrightarrow{p} \theta^* = \theta_0$, and we have $E[R\{Y-m(X;\beta^*)\}|\pi(X;\theta_0)] = E\{R|\pi(X;\theta_0)\}E\{Y-m(X;\beta^*)|\pi(X;\theta_0)\}$ due to that

$R \perp\!\!\!\perp Y | \pi(X; \theta_0)$ and $R \perp\!\!\!\perp X | \pi(X; \theta_0)$. It can be checked that

$$\begin{aligned}
&\frac{d}{d\beta}|_{\beta=\beta^*} E\Big[\frac{R\{Y - m(X;\beta)\}}{E\{R|\pi(X;\theta_0)\}} + m(X;\beta)\Big] \\
&= \frac{d}{d\beta}|_{\beta=\beta^*} \Bigg\{ E\Big[E\Big\{\frac{R\{Y - m(X;\beta)\}}{E\{R|\pi(X;\theta_0)\}}|\pi(X;\theta_0)\Big\}\Big] + E\{m(X;\beta)\}\Bigg\} \\
&= \frac{d}{d\beta}|_{\beta=\beta^*} \Big[E\{Y - m(X;\beta)\} + E\{m(X;\beta)\}\Big] = 0, \\
&n^{\frac{1}{2}}(\widehat{\mu} - \mu) = n^{\frac{1}{2}}(P_n - P)\Big\{\frac{RY}{\pi(X;\theta_0)} \\
&- \frac{R - \pi(X;\theta_0)}{\pi(X;\theta_0)}\big[E\{Y - m(X;\beta^*)|\pi(X;\theta_0)\} + m(X;\beta^*)\big] - \mu\Big\} \\
&+ \frac{d}{d\theta}|_{\theta=\theta_0} E\Big[\frac{R\{Y - m(X;\beta^*)\}}{E\{R|\pi(X;\theta)\}}\Big] n^{\frac{1}{2}}(\widehat{\theta} - \theta_0) + o_p(1).
\end{aligned}$$

We can also check that, if $m(X;\beta)$ is the correct model for $Y|X$, then $m(X;\widehat{\beta}) \xrightarrow{p} m(X;\beta_0) = E(Y|X)$. Then

$$\begin{aligned}
&\frac{d}{d\theta}|_{\theta=\theta^*} E\Big[\frac{R\{Y - m(X;\beta^*)\}}{E\{R|\pi(X;\theta)\}}\Big] = \frac{d}{d\theta}|_{\theta=\theta^*} E\Big[\frac{R\{Y - m(X;\beta^*)\}}{E\{R|\pi(X;\theta)\}}|X\Big] \\
&= \frac{d}{d\theta}|_{\theta=\theta^*} E\Big[\frac{E(R|X)\{E(Y|X) - m(X;\beta_0)\}}{E\{R|\pi(X;\theta)\}}|X\Big] = 0,
\end{aligned}$$

and $n^{\frac{1}{2}}(\widehat{\mu} - \mu) =$

$$\begin{aligned}
&n^{\frac{1}{2}}(P_n - P)\Big[\frac{RY}{E\{R|\pi(X;\theta^*)\}} - \frac{R - E\{R|\pi(X;\theta^*)\}}{E\{R|\pi(X;\theta^*\}} m(X;\beta_0) - \mu\Big] \\
&+ \frac{d}{d\beta}|_{\beta=\beta_0} E\Big[\frac{R\{Y - m(X;\beta)\}}{E\{R|\pi(X;\theta^*)\}} + m(X;\beta)\Big] n^{\frac{1}{2}}(\widehat{\beta} - \beta_0) + o_p(1).
\end{aligned}$$

When both working models are correct, then we have

$$n^{\frac{1}{2}}(\widehat{\mu} - \mu) = n^{\frac{1}{2}}(P_n - P)\Big\{\frac{RY}{\pi(X;\theta_0)} - \frac{R - \pi(X;\theta_0)}{\pi(X;\theta_0)} m(X;\beta_0) - \mu\Big\} + o_p(1),$$

because $E\{R|\pi(X;\theta_0)\} = \pi(X;\theta_0)$.

Appendix C: Weight Trimming:

We implemented the implemented the usual AIPW DR estimator after trimming the estimated propensity score at and smaller than 0.1. Our simulation results show that weight trimming reduces impact of large weight and mitigates the problem of "disastrous" behavior in simulation scenario 1. However, it may also introduce unignorable bias and lead to lower coverage probability than the nominal level in some scenarios. For example, in simulation scenario 2, when the outcome model is incorrect but the propensity score model is correct, the bias and coverage probability are 0.11 and 0.90 respectively for $n = 200$ and are 0.077 and 0.85 for $n = 1000$. Performance of the trimming method heavily depends on specifics of the scenarios. One possible reason for the difference in behaviors of the trimming method between settings 1 and 2 is that in setting 1 the outcome model is only mildly misspecified but in setting 2 the outcome model is more severely misspecified.

References

Tsiatis, A. A. (2006). Semiparametric Theory and Missing Data. New York: Springer.