TMLE of Primitive of Cumulative Distribution of Conditional Treatment Effect

Mark van der Laan
Division of Biostatistics, University of California, Berkeley
laan@berkeley.edu

June 30, 2016

Abstract

Keywords:

1 Introduction

Let $O = (W, A, Y) \sim P_0$ and let the model $\mathcal{M}$ be nonparametric. Let $B(P)(W) = E_P(Y \mid A = 1, W) - E_P(Y \mid A = 0, W)$ and let $\Psi_y(P) = \int_{-1}^{y} E_P I(B(P)(W) > x)dx = \int_{-1}^{y} S_B(x)dx$ be the target parameter of interest, where $S_B(x) = P(B(W) > x)$. We note that we can also write this target parameter as follows:

$$\Psi_y(P) = E_P(\min(B(P)(W), y) + 1) = 1 + E_P \min(B(P)(W), y).$$

The latter representation demonstrates that this parameter is smoother in P than $S_B(x) = E_P I(B(P)(W) > x)$: instead of having B in an indicator it is now truncated from above (i.e. from non differentiable to continuous with a point of non differentiability). We will prove that Ψ_y is pathwise differentiable at P with efficient influence curve given by:

$$D_y^*(P) = \frac{2A-1}{g(A|W)} I(B(P)(W) \leq y)(Y - \bar{Q}(A, W)) + \int_{-1}^{y} I(B(P)(W) > x) - \Psi_y(P).$$

Subsequently, we will study the remainder $P_0 D_y^*(P) + \Psi_y(P) - \Psi_y(P_0)$.

1.1 Pathwise differentiability of Ψ_y

We first note that $b \rightarrow I(b > x)$ is a function in b that jumps from 0 to 1 at x. One can think of this as a cumulative distribution function of a pointmass 1 at x. That is, we can define $F_{0,x}(b) = I(b > x) = \int_{-1}^{b} dF_{0,x}(u)$, where $dF_{0,x}$

is the degenerate probability distribution/density that puts mass 1 at x. Thus $\Psi_y(P) = \int_{-1}^{y} E_P F_{0,x}(B(P)(W))dx$. This suggest approximating $F_{0,x}$ with $F_{h,x}(b) = \int_{-1}^{b} K((u-x)/h)/h du$, where K is a mean zero density with support $[-1,1]$. Note that $F_{h,x}(b)$ approximates the step function $F_{0,x}(b)$ by smoothing its jump at x from 0 to 1 over a little interval $[x-h, x+h]$. This suggest defining an approximate parameter

$$\Psi_{y,h}(P) = \int_{-1}^{y} E_P F_{h,x}(B(P)(W))dx.$$

Let's consider the function $\Psi_{y,h}(B) = \int_{-1}^{y} E_P F_{h,x}(B(W))dx$, treating P as known in the expectation E_P, and let's aim to understand if this parameter is pathwise differentiable and its behavior at $h \to 0$. We have

$$\begin{aligned} \Psi_{y,h}(B_\epsilon) - \Psi_{y,h}(B) &= \int_{-1}^{y} E_P\{F_{h,x}(B_\epsilon(W)) - F_{h,x}(B(W))\}dx \\ &= \int_{-1}^{y} E_P \frac{d}{dB(W)} F_{h,x}(B(W))(B_\epsilon(W) - B(W))dx \\ &\quad + \int_{-1}^{y} E_P \frac{d^2}{db(W)} F_{h,x}(b(W))(B_\epsilon(W) - B(W))^2/2, \end{aligned}$$

for a $b = b(x, h, W, \epsilon)$ in the interval spanned by the values $B_\epsilon(W)$ and $B(W)$. Let's denote this last second order term with $R_{2,h,y}(B_\epsilon, B)$. We have

$$\frac{d}{db} F_{h,x}(b) = \frac{d}{db} \int_{-1}^{b} K((u-x)/h)/h du = K((b-x)/h)/h.$$

Thus, we obtain

$$\begin{aligned} \Psi_{y,h}(B_\epsilon) - \Psi_{y,h}(B) &= E_P \int_{-1}^{y} K((B(W)-x)/h)/h dx (B_\epsilon(W) - B(W)) \\ &\quad + R_{2,h,y}(B_\epsilon, B) \\ &= E_P F_{h,B(W)}(y)(B_\epsilon(W) - B(W)) + R_{2,h,y}(B_\epsilon, B) \\ &= E_P(F_{0,B(W)}(y)(B_\epsilon(W) - B(W)) + R_{2,h,y}(B_\epsilon, B) \\ &\quad + E_P(F_{h,B(W)}(y) - F_{0,B(W)}(y))(B_\epsilon(W) - B(W)). \end{aligned}$$

Note $F_{0,B(W)}(y) = I(B(W) \leq y)$. As $h \to 0$, we have that $\| F_{h,B(W)}(y) - F_{0,B(W)}(y) \|_P = O(h)$ converges to zero. Using Cauchy-Schwarz inequality, the last term is bounded by $\| F_{h,B(W)}(y) - F_{0,B(W)}(y) \|_P \| B_\epsilon - B \|_P$, which is thus $O(h \| B_\epsilon - B \|_P)$. We also note that the second derivative of $F_{h,x}$ in $R_{2,h,y}()$ behaves as $1/h$ so that

$$R_{2,h,y}(B_\epsilon, B) = O(\| B_\epsilon - B \|^2 /h).$$

Thus,

$$\Psi_{y,h}(B_\epsilon) - \Psi_{y,h}(B) = E_P I(B(W) \leq y)(B_\epsilon(W) - B(W)) + O(h \| B_\epsilon - B \|_P) + O(\| B_\epsilon - B \|_P^2 /h).$$

Note $B_\epsilon - B = O(\epsilon)$. Given a path $\{P_\epsilon : \epsilon\}$, we can select $h = h(\epsilon) \to 0$ so that $\epsilon/h \to 0$ so that

$$\frac{\Psi_{y,h_\epsilon}(B(P_\epsilon)) - \Psi_{y,h_\epsilon}(B(P))}{\epsilon} = E_P I(B(W) \leq y)(B_\epsilon(W) - B(W))/\epsilon + o(1).$$

The right-hand side will converge to

$$PD^*_{y,1}(P)S,$$

where S is the score of $\{P_\epsilon : \epsilon\}$ and

$$D^*_{y,1}(P)(O) = \frac{2A-1}{g(A \mid W)} I(B(W) \leq y)(Y - \bar{Q}(A,W)).$$

In this way, we have proven that for such a sequence $h = h(\epsilon)$

$$\lim_{\epsilon \to 0} \frac{\Psi_{h_\epsilon}(P_\epsilon) - \Psi_{h_\epsilon}(P)}{\epsilon} = PD^*_y(P)S,$$

where

$$D^*_y(P) = D^*_{y,1}(P)(O) + D^*_{y,2}(P)(O),$$

and

$$D^*_{y,2}(P)(O) = \int_{-1}^{y} I(B(P)(W) > x)dx - \Psi_y(P).$$

So we have the following theorem:

Theorem 1 *Consider the target parameter $\Psi_{y,h} : \mathcal{M} \to \mathbb{R}$ defined by*

$$\Psi_{y,h}(P) = \int_{-1}^{y} E_P F_{h,x}(B(P)(W))dx.$$

We have that for a path $\{P_\epsilon : \epsilon\}$ with score S at $\epsilon = 0$

$$\frac{\Psi_{y,h}(P_\epsilon) - \Psi_{y,h}(P)}{\epsilon} = PD^*_{y,h}(P)S + O(\epsilon/h),$$

where

$$D^*_{y,h}(P) = \frac{2A-1}{g(A \mid W)} F_{h,B(W)}(y)(Y - \bar{Q}(A,W)) + \int_{-1}^{y} F_{h,x}(B(P)(W))dx - \Psi_{y,h}(P).$$

For $h \to 0$ and $\epsilon/h \to 0$, we have

$$\lim_{\epsilon \to 0} \frac{\Psi_{y,h_\epsilon}(P_\epsilon) - \Psi_{y,h_\epsilon}(P)}{\epsilon} = PD^*_y(P)S.$$

In order to show that this result implies pathwise differentiability of $\Psi_{y,0}$, we would need to show the following lemma.

Lemma 1

$$\begin{aligned}
&\frac{\Psi_{y,h_\epsilon}(P_\epsilon)-\Psi_{y,0}(P_\epsilon)-\{\Psi_{y,h_\epsilon}(P)-\Psi_{y,0}(P)\}}{\epsilon} \\
&= \frac{\int_{-1}^{y} P(F_{h,x}-F_{0,x})(B(P_\epsilon)dx-\int_{-1}^{y} P(F_{h,x}-F_{0,x})(B(P))dx}{\epsilon} \\
&\to 0,
\end{aligned}$$

as $\epsilon \to 0$, $h_\epsilon \to 0$, $\epsilon/h_\epsilon \to 0$.

We have to check if the latter is true, but I suspect it is since we know that $(\Psi_{y,h_\epsilon}(P_\epsilon)-\Psi_{y,h_\epsilon}(P))/\epsilon$ converges nicely (no singularity), and $(F_{h,x}-F_{0,x})(B(W))$ converges to zero in $L^2(P)$-norm, which should all that is relevant since we are averaging over distribution of W. In that case, we have the following theorem.

Theorem 2 *Recall* $F_{0,x}(b)=I(b>x)$. *Consider the target parameter* $\Psi_y : \mathcal{M} \to \mathbb{R}$ *defined by*

$$\Psi_y(P)=\int_{-1}^{y} E_P F_{0,x}(B(P)(W))dx.$$

We have that for a path $\{P_\epsilon : \epsilon\}$ *with score* S *at* $\epsilon = 0$

$$\lim_{\epsilon\to 0}\frac{\Psi_y(P_\epsilon)-\Psi_y(P)}{\epsilon}=PD_y^*(P)S,$$

where

$$D_y^*(P)=\frac{2A-1}{g(A\mid W)}F_{0,B(W)}(y)(Y-\bar{Q}(A,W))+\int_{-1}^{y} F_{0,x}(B(P)(W))dx-\Psi_y(P).$$

1.2 Study of Remainder

Let's suppress y in the notation for now. Let's study the remainder

$$R_{20}(P,P_0)=P_0D^*(P)+\Psi(P)-\Psi(P_0).$$

Let $Q_W(P) = Q_W(P_0)$, since the contribution from $Q_W(P), Q_W(P_0)$ is trivial. We have

$$
\begin{aligned}
P_0 D^*(P) + \Psi(P) - \Psi(P_0) &= E_{P_0} \frac{2A-1}{g(A \mid W)} I(B(W) < y)(Y - \bar{Q}(A,W)) \\
&\quad + \int_{-1}^{y} I(B(W) > x)dx - \int_{-1}^{y} P_0 I(B > x)dx \\
&\quad + \int_{-1}^{y} P_0 I(B > x)dx - \int_{-1}^{y} P_0 I(B_0 > x)dx \\
&= E_{P_0} \frac{2A-1}{g(A \mid W)} I(B(W) < y)(Y - \bar{Q}(A,W)) \\
&\quad + \int_{-1}^{y} I(B(W) > x)dx - \int_{-1}^{y} P_0 I(B_0 > x)dx \\
&= E_{P_0} \frac{g_0(1 \mid W)}{g(1 \mid W)} I(B(W) < y)(\bar{Q}_0 - \bar{Q})(1,W) \\
&\quad - E_{P_0} \frac{g_0(0 \mid W)}{g(0 \mid W)} I(B(W) < y)(\bar{Q}_0 - \bar{Q})(0,W) \\
&\quad + \int_{-1}^{y} I(B(W) > x)dx - \int_{-1}^{y} P_0 I(B_0 > x)dx \\
&= E_{P_0} I(B(W) < y)(B_0 - B)(W) + \int_{-1}^{y} P_0 I(B > x)dx \\
&\quad - \int_{-1}^{y} P_0 I(B_0 > x)dx + R_{20,1}(g, g_0, \bar{Q}, \bar{Q}_0),
\end{aligned}
$$

where

$$
\begin{aligned}
R_{20,1}(g, g_0, \bar{Q}, \bar{Q}_0) &= E_{P_0} \left(\frac{g_0(1 \mid W)}{g(1 \mid W)} - 1 \right) I(B(W) < y)(\bar{Q}_0 - \bar{Q})(1,W) \\
&\quad - E_{P_0} \left(\frac{g_0(0 \mid W)}{g(0 \mid W)} - 1 \right) I(B(W) < y)(\bar{Q}_0 - \bar{Q})(0,W).
\end{aligned}
$$

So it remains to analyze:

$$
-R_{20,2}(B, B_0) \equiv \int_{-1}^{y} P_0\{I(B > x) - I(B_0 > x)\}dx - P_0 I(B < y)(B - B_0),
$$

and note that $R_{20}(P,P_0) = R_{20,1}(P,P_0) + R_{20,2}(P,P_0)$. By Fubini's theorem, we can write this as:

$$
\begin{aligned}
&R_{20,2}(B,B_0) = E_{P_0}\int_{-1}^{y} I(B(W) > x)dx - E_{P_0}\int_{-1}^{y} I(B_0(W) > x)dx - P_0 I(B < y)(B - B_0)\\
&= E_{P_0}(\min(y,B(W)) + 1) - E_{P_0}(\min(y,B_0(W)) + 1) - P_0 I(B < y)(B - B_0)\\
&= E_{P_0}(\min(y,B(W)) - \min(y,B_0(W))) - P_0 I(B < y)(B - B_0)\\
&= E_{P_0} I(B < y, B_0 < y)(B - B_0) + E_{P_0} I(B < y, B_0 > y)(B - y) + E_{P_0} I(B > y, B_0 < y)(y - B_0)\\
&-P_0 I(B < y)(B - B_0)\\
&= P_0\{I(B < y, B_0 < y) - I(B < y)\}(B - B_0) + P_0 I(B < y, B_0 > y)(B - y)\\
&+P_0 I(B > y, B_0 < y)(y - B_0)\\
&= P_0 I(B < y, B_0 > y)(B_0 - B) + P_0 I(B < y, B_0 > y)(B - y)\\
&+P_0 I(B > y, B_0 < y)(y - B_0)\\
&= P_0 I(B < y, B_0 > y)(B_0 - y) + P_0 I(B > y, B_0 < y)(y - B_0)
\end{aligned}
$$

where we used that $I(B < y, B_0 < y) - I(B < y) = I(B < y)(I(B_0 < y) - 1) = -I(B < y, B_0 > y)$. Note that both terms are positive. Note that

$$
\begin{aligned}
0 \leq P_0 I(B < y, B_0 > y)(B_0 - y) &\leq P_0 I(B < y, B_0 > y)(B_0 - B)\\
0 \leq P_0 I(B > y, B_0 < y)(y - B_0) &\leq P_0 I(B > y, B_0 < y)(B - B_0)
\end{aligned}
$$

So we have that $R_{20,2}(B,B_0) \geq 0$ and

$$
R_{20,2}(B,B_0) \leq P_0 I(B < y, B_0 > y)(B_0 - B) + P_0 I(B > y, B_0 < y)(B - B_0).
$$

This is indeed a higher order remainder. For example, using Cauchy-Schwarz inequality, we could bound $R_{20,2}$ as follows:

$$
\begin{aligned}
\mid R_{20,2}(B,B_0) \mid &\leq \parallel B - B_0 \parallel_{P_0} \parallel I(B < y, B_0 > y) \parallel_{P_0}\\
&\quad + \parallel B - B_0 \parallel_{P_0} \parallel I(B > y, B_0 < y) \parallel_{P_0}\\
&= \parallel B - B_0 \parallel_{P_0} \left\{ \sqrt{P_0(B < y, B_0 > y)} + \sqrt{P(B > y, B_0 < y)} \right\}.
\end{aligned}
$$

Using this Cauchy-Schwarz bounding, the remainder appears to behave as a $(B - B_0)^{3/2}$, intuitively speaking. However, one can also bound it as

$$
R_{20,2}(B,B_0) \leq \parallel B_0 - B \parallel_\infty \{P_0(B < y, B_0 > y) + P_0(B > y, B_0 < y)\},
$$

which is a second order difference. To further understand this remainder, we note that

$$
\begin{aligned}
P_0(B < y, B_0 > y) &= E_0 I(B(W) < y, B_0(W) > y)\\
&= E_0 I(B(W) < y, B_0(W) > y) I(B_0(W) - B(W) > B_0(W) - y)\\
&\leq E_0 I(B(W) < y, B_0(W) > y) I(\parallel B_0 - B \parallel_\infty > B_0(W) - y)\\
&= P_0(y < B_0(W) < y + \parallel B_0 - B \parallel_\infty, B(W) < y)\\
&\leq F_{B_0}(y + \parallel B_0 - B \parallel_\infty) - F_{B_0}(y),
\end{aligned}
$$

where $F_{B_0}(y) = P_0(B_0 \leq y)$ is the cdc of B_0. If we assume that F_{B_0} is Lipsnitz-continuous, then this can be bounded by a constant times $\parallel B_0 - B \parallel_\infty$. Similarly,

$$
\begin{aligned}
P_0(B > y, B_0 < y) &= E_0 I(B(W) > y, B_0(W) < y) \\
&= E_0 I(B(W) > y, B_0(W) < y) I(y - B_0(W) < B(W) - B_0(W)) \\
&\leq E_0 I(B(W) > y, B_0(W) < y) I(y - B_0(W) < \parallel B - B_0 \parallel_\infty) \\
&= P_0(y - \parallel B - B_0 \parallel_\infty < B_0(W) < y, B(W) > y) \\
&\leq F_{B_0}(y) - F_{B_0}(y - \parallel B_0 - B \parallel_\infty).
\end{aligned}
$$

This proves the following theorem.

Theorem 3 *Let*

$$R_{20}(P, P_0) = P_0 D^*(P) + \Psi(P) - \Psi(P_0).$$

We have

$$R_{20}(P, P_0) = R_{20,1}(g, g_0, \bar{Q}, \bar{Q}_0) + R_{20,2}(B, B_0),$$

where

$$
\begin{aligned}
R_{20,1}(g, g_0, \bar{Q}, \bar{Q}_0) &= E_{P_0} \left(\frac{g_0(1 \mid W)}{g(1 \mid W)} - 1 \right) I(B(W) < y)(\bar{Q}_0 - \bar{Q})(1, W) \\
&\quad - E_{P_0} \left(\frac{g_0(0 \mid W)}{g(0 \mid W)} - 1 \right) I(B(W) < y)(\bar{Q}_0 - \bar{Q})(0, W),
\end{aligned}
$$

and

$$0 \leq R_{20,2}(B, B_0) = P_0 I(B < y, B_0 > y)(B_0 - B) + P_0 I(B > y, B_0 < y)(B - B_0).$$

We can bound the latter term in the following two ways.

$$
\begin{aligned}
R_{20,2}(B, B_0) &\leq \parallel B - B_0 \parallel_{P_0} \left\{ \sqrt{P_0(B < y, B_0 > y)} + \sqrt{P(B > y, B_0 < y)} \right\} \\
R_{20,2}(B, B_0) &\leq \parallel B_0 - B \parallel_\infty \left\{ P_0(B < y, B_0 > y) + P_0(B > y, B_0 < y) \right\},
\end{aligned}
$$

and

$$
\begin{aligned}
P_0(B < y, B_0 > y) &\leq F_{B_0}(y + \parallel B_0 - B \parallel_\infty) - F_{B_0}(y) \\
P_0(B > y, B_0 < y) &\leq F_{B_0}(y) - F_{B_0}(y - \parallel B_0 - B \parallel_\infty),
\end{aligned}
$$

where F_{B_0} is the cumulative distribution function of B_0.