# Ensemble Super Learning of the Optimal Treatment Rule for a Continuous Treatment, and Inference for its Counterfactual Mean Outcome

Mark van der Laan[1], Alex Luedtke[2]
University of California, Berkeley[1]
Fred Hutchinson Cancer Research Institute[2]
laan@berkeley.edu

September 15, 2016

**Abstract**

We consider the problem that we observe $n$ independent and identically distributed observations of a random variable consisting of baseline covariates, a subsequent continuous treatment, and a final outcome of interest. For a given rule for assigning treatment, we define the mean outcome under the distribution of the data in which the conditional distribution of treatment, given covariates, is replaced by this rule. Under the randomization assumption and a positivity assumption this equals the counterfactual mean outcome under this rule. The optimal treatment rule is defined as the rule minimizing the rule-specific counterfactual mean outcome over all rules. For an estimator of the optimal rule, we define its measure of performance as the counterfactual mean outcome under the estimate of the optimal rule when applied to a training sample, averaged across training samples. We refer to this as the risk of the estimator. The mean outcome under a candidate treatment rule is not a pathwise differentiable target parameter, so that estimation of the risk is a non-regular problem. We define a cross-validated targeted maximum likelihood estimator (CV-TMLE) of this risk of the estimator, which depends on an amount of regularization defined by a smoothing bandwidth. We analyze this CV-TMLE and establish that it is asymptotically normally distributed under weak regularity conditions, and we propose a data adaptive selector of the bandwidth that guarantees that the CV-TMLE converges to the target at the optimal unknown rate depending on the underlying unknown smoothness of the outcome regression as a function of treatment. This allows us to obtain efficient confidence intervals for the risk of the estimator. Subsequently, we propose an ensemble super-learner of the optimal rule that uses cross-validation to select the minimal risk estimator among a library of candidate estimators of the optimal rule. We prove a finite sample oracle inequalities and establishing that it will perform asymptotically as well as the oracle selector that selects the best algorithm in the library of candidate estimators, under specified realistic conditions. We propose and analyze a particular type of $\delta$-net estimator to be included in the library and show that it guarantees good rates of convergence of the super-learner.

**Keywords**: Asymptotically linear estimator, causal inference, continuous treatment, counterfactual, cross-validation, efficient influence curve, ensemble learning, individualized treatment, influence curve, super-learning, targeted minimum loss-based estimation

# 1 Introduction

Observational and randomized studies can be utilized to learn the optimal individualized treatment rule that minimizes the counterfactual mean outcome of interest such as a failure rate. This article focusses on the case that the treatment of interest is only assigned at one point in time, so that the goal is to learn the optimal treatment allocation rule for that single time point. In the case that the study is a longitudinal study observing treatment decisions across multiple time points, then it is possible to learn the optimal dynamic treatment regimen across multiple time points, where each time-specific rule can be a function of the observed past.

Most of the literature on optimal dynamic treatments focussed on the case that the treatment is binary or more generally discrete valued. A recent article Kosorok et al. proposed a machine learning algorithm for learning the optimal dynamic treatment for a continuous valued treatment.

# 2 Formulation of the Statistical Estimation Problem

**Observed data:** We observe $n$ independent and identically distributed (i.i.d.)

observations on an random variable $O = (X, A, Y) \sim P_0$, where $X$ is a vector of baseline covariates, $A$ is a continuous valued treatment drawn after $X$, and $Y$ is the final outcome. We make the convention that low values of $Y$ represent good (e.g., patient) outcomes. We consider the case that $Y \in \{0, 1\}$ or that it is continuous with values in $(0, 1)$. Let $A \in [0, 1]$ and $X \in [0, \tau] \subset \mathbb{R}^d$. Let $P_n$ be the empirical probability distribution of $O_1, \ldots, O_n$.

**Statistical Model:** For a given $P$, $G = G(P)$ is the conditional distribution of $A$, given $X$, $g = g(P)$ is its conditional density w.r.t. Lebesgue measure, $\bar{Q}(P)(a, x) = E_P(Y \mid A = a, X = x)$ is the outcome regression, and $Q_X = Q_X(P)$ is the probability distribution of $X$. We will assume a strong positivity assumption

$$\inf_a g_0(a \mid X) > \delta > 0 \text{ a.e. for some } \delta > 0. \tag{1}$$

Let $\mathcal{G}$ be an assumed parameter space for $g_0$, which includes this positivity assumption. In addition, we assume that $a \to \bar{Q}_0(a, X)$ is $J$-times continuously differentiable with a $J$-th derivative $\bar{Q}_0^{(J)}(a, X)$ that is uniformly bounded away from infinity, while we leave $Q_{0,X}$ unspecified. Let $Q = (\bar{Q}, Q_X)$. Let $\mathcal{M}(J) = \{P : g(P) \in \mathcal{G}, \parallel \bar{Q}^{(J)} \parallel_\infty < \infty\}$ be the resulting statistical model for $P_0$. Let $f(P)(x) \equiv \arg\min_a \bar{Q}(P)(a, x)$ be the optimal treatment rule maximizing $f \to E_P \bar{Q}_0(f(X), X)$, where the latter equals the counterfactual mean $E_P Y_f$ under a causal model and randomization assumption. Consider the following assumption:

$$\frac{d}{df(P)(x)} \bar{Q}_0(f(P)(x), x) = 0 \text{ for } P\text{-a.e. } x. \tag{2}$$

In addition, consider the assumption

$$\inf_{a,x} \bar{Q}_0^{(2)}(a, x) > 0 \tag{3}$$

Let $\mathcal{M}_I(J) \equiv \{P \in \mathcal{M}(J) : P \text{ satisfies } (2) and (3)\}$ be the statistical model that also assumes (2) and (3). The model $\mathcal{M}_I(J)$ will only be used in the Appendix to establish an improved oracle inequality for the case that these assumptions hold.

**Target parameters: risk of candidate estimator and optimal treatment rule:** For a given treatment allocation rule $f(X)$, let

$$V_0(f) = E_0 \bar{Q}_0(f(X), X)$$

be the measure of its performance, and we will refer to $V_0(f)$ as the risk of $f$. Let $B_n \in \{0, 1\}^n$ be a cross-validation scheme, and $P^1_{n,B_n}, P^0_{n,B_n}$ are the empirical probability distributions of the validation sample $\{i : B_n(i) = 1\}$

and training sample $\{i : B_n(i) = 0\}$, respectively. Throughout, we assume that $EB_n = p > 0$ for some $p > 0$ and that $B_n$ has only a finite number of realizations, such as in $V$-fold cross-validation. For a given candidate estimator $\hat{f}$ we define its risk as

$$V_0(\hat{f}) \equiv E_{B_n} V_0(\hat{f}(P^0_{n,B_n})).$$

Our first goal is to provide statistical inference for this data adaptive target parameter $V_0(\hat{f})$. This includes as special case statistical inference for $V_0(f)$ for any given rule $f$. Since $V_0(f)$ is not pathwise differentiable this represents a non-regular estimation problem. The optimal rule $f_0(x) \equiv \arg\min_a \bar{Q}_0(a, X)$ minimizes the risk function $V_0(f)$ over all rules $f$: $V_0(f_0) = \min_f V_0(f)$. Our second goal is to construct a ensemble super-learner of $f_0$. Given a library of candidate estimators of $f_0$, we will define the super-learner as the estimator that minimizes our estimator of $V_0(\hat{f})$.

## 2.1 Organization of article

In Section 3 we present a pathwise differentiable $\epsilon$-approximation $V_0(\hat{f}, \epsilon)$ of $V_0(\hat{f})$ relying on a kernel smooth with bandwidth $\epsilon$. In this section we present a double robust asymptotically efficient cross-validated TMLE of $V_0(\hat{f}, \epsilon)$. Without making smoothness conditions, we prove that this CV-TMLE converges at rate $(n\epsilon_n)^{-0.5}$ to $V_0(\hat{f}, \epsilon_n)$ with a mean zero normal limit distribution with specified variance, along general sequences $\epsilon_n \to 0$. Subsequently, we prove that under smoothness conditions, the CV-TMLE converges at rate $n^{-J_0/(2J_0+1)}$ to $V_0(\hat{f})$ with a normal limit distribution with a specified mean and same variance, where $J_0$ is the underlying smoothness of $\bar{Q}_0$. This result relies on $\epsilon_n$ to behave as $n^{-1/(2J_0+1)}$. In Section 4 we propose a data adaptive selector of $\epsilon$ that achieves this optimal known rate. These results allow us to construct a 0.95-confidence interval for $V_0(\hat{f})$. In Section 5 we define the super-learner based on this CV-TMLE estimate of the risk $V_0(\hat{f}, \epsilon_n)$ of a candidate estimator. We also present super-learner based on CV-IPTW and CV-AIPTW estimators of $V_0(\hat{f}, \epsilon_n)$. In Section 6 we establish an oracle inequality for this super-learner showing that it is asymptotically equivalent with an oracle selector. Firstly we present an oracle inequality measuring performance w.r.t $V_0(\hat{f}, \epsilon_n)$ and subsequently, under smoothness conditions, we present the oracle inequality measuring performance w.r.t. $V_0(\hat{f})$. We show that the rate at which $\epsilon_n$ should converge to zero to get the best oracle inequality corresponds with the optimal rate $n^{-1/(2J_0+1)}$ above, so that the same $\epsilon_n$-selector can be used. In our presentation of the oracle inequality we focus on the case $G_0$ is known, but generalizations are immediately available, but now involving an additional term of the same order as the rate at which $G_n$ converges to $G_0$. In Section 7 we present a $\mathcal{F}$-specific $\delta$-net estimator of the optimal rule $f_0$ and prove that it converges at a specified rate implied by the entropy of the parameter space $\mathcal{F}$ of $f_0$. Incorporation of this estimator in the super-learner library guarantees this rate for the super-learner. We conclude with a discussion in Section 8. Some additional technical results are presented in the Appendix.

# 3 Statistical Inference for risk of candidate estimator of optimal rule

## 3.1 Approximating risk of candidate estimator with pathwise differentiable risk

Define the following approximation of $V_0(f)$:

$$V_P(f,\epsilon) \equiv E_P \int \frac{1}{\epsilon} k\left(\frac{a - f(X)}{\epsilon}\right) \bar{Q}(a,X)da,$$

where $k$ is a $J-1$-orthogonal univariate kernel satisfying $\int k(x)dx = 1$, $\int k(x)x^j dx = 0$, $j = 1, \ldots, J-1$, and $k$ has support $[-1,1]$. For the sake of notation, let $k_{f(X),\epsilon}(a) \equiv \epsilon^{-1}k((a - f(X))/\epsilon)$. Note that $V_P(f,\epsilon) = EY_{f_\epsilon}$ is the counterfactual mean under a stochastic intervention $f_\epsilon$ (i.e., conditional distribution of $A$, given $X$) defined by drawing $A$ from $k_{f(X),\epsilon}$. Clearly this stochastic intervention approximates the deterministic intervention $f$.

For a given $\epsilon$, $V_P(f,\epsilon)$ is a pathwise differentiable target parameter with efficient influence curve $D^*_{f,\epsilon}(P)$ given by:

$$D^*_{f,\epsilon}(P) = \frac{k_{f(X),\epsilon}(A)}{g(A \mid X)}(Y - \bar{Q}(A,X)) + \int k_{f(X),\epsilon}(a)\bar{Q}(a,X)da - V_P(f,\epsilon).$$

We have the following first order expansion for this target parameter:

$$V_P(f,\epsilon) - V_{P_0}(f,\epsilon) = -P_0 D^*_{f,\epsilon}(P) + R_{20,f,\epsilon}(P,P_0),$$

where

$$R_{20,f,\epsilon}(P,P_0) = E_{P_0} \int_a k_{f(X),\epsilon}(a)\frac{(g-g_0)(a \mid X)}{gg_0(a \mid X)}(\bar{Q} - \bar{Q}_0)(a,X)da$$

is a second order remainder. We will also use the notation $D^*_{f,\epsilon}(Q,G)$ and $R_{20,f,\epsilon}(Q,G,Q_0,G_0)$ for $D^*_{f,\epsilon}(P)$ and $R_{20,f,\epsilon}(P,P_0)$, respectively.

Let $\hat{f}$ be a candidate estimator of $f_0$ so that $f_n = \hat{f}(P_n)$ is the estimated rule. Let $V_0(\hat{f}, \epsilon) \equiv E_{B_n} V_0(\hat{f}(P^0_{n,B_n}), \epsilon)$, and note that this reduces to $V_0(f, \epsilon)$ if $\hat{f} = f$.

## 3.2 Defining the CV-TMLE of the $\epsilon$-risk of a candidate estimator

The candidate estimator $\hat{f}$ is given. Let $\bar{Q}_n = \hat{\bar{Q}}(P_n)$ and $G_n = \hat{G}(P_n)$ be initial estimators of $\bar{Q}_0$ and $G_0$, respectively. In addition, let $\bar{Q}_{n,B_n}, G_{n,B_n}, f_{n,B_n} = \hat{f}(P^0_{n,B_n})$ be the estimators applied to the training sample. Let $\text{Logit}\bar{Q}_{n,B_n,\delta} = \text{Logit}\bar{Q}_{n,B_n} + \delta C(G_{n,B_n}, f_{n,B_n}, \epsilon)$ be the least favorable submodel through $\bar{Q}_{n,B_n}$ at $\delta = 0$, where $C(G, f, \epsilon) = k_{f(X),\epsilon}(A)/g(A|X)$. Let $\epsilon_n$ be the maximizer of the cross-validated log-likelihood:

$$\delta_n = \arg\min_{\delta} E_{B_n} P^1_{n,B_n} L(\bar{Q}_{n,B_n,\epsilon}),$$

where

$$L(\bar{Q})(O) = -\left\{ Y \log \bar{Q}(A, X) + (1 - Y) \log(1 - \bar{Q}(A, X)) \right\}.$$

This defines the $B_n$-specific TMLE $\bar{Q}^*_{n,B_n} = \bar{Q}_{n,B_n,\delta_n}$ of $V_0(f_{n,B_n}, \epsilon_n)$ (treating $f_{n,B_n}$ as fixed). We now define the CV-TMLE of $V_0(\hat{f}, \epsilon) \equiv E_{B_n} V_0(f_{n,B_n}, \epsilon)$ as

$$V^*_n(\hat{f}, \epsilon) \equiv E_{B_n} E_{P^1_{n,B_n}} \int k_{f_{n,B_n}(X),\epsilon}(a) \bar{Q}^*_{n,B_n}(a, X) da.$$

We note that this CV-TMLE solves the cross-validated efficient influence curve equation:

$$0 = E_{B_n} P^1_{n,B_n} D^*_{f_{n,B_n},\epsilon}(Q^*_{n,B_n}, G_{n,B_n}),$$

but where $V_P(f_{n,B_n}, \epsilon)$ in $D^*_{f_{n,B_n},\epsilon_n}$ is estimated with $V^*_n(\hat{f}, \epsilon)$. For the sake of notational convenience, we suppress this in the notation.

## 3.3 Asymptotic normality of CV-TMLE as estimator of $\epsilon$-risk of candidate estimator

The following theorem establishes asymptotic normality of the CV-TMLE $V^*_n(\hat{f}, \epsilon)$ of $V_0(\hat{f}, \epsilon)$ without making smoothness assumptions on $\bar{Q}_0$.

**Theorem 1** *Let $P_0 \in \mathcal{M}(J = 0)$. Define the following $L^2(P_0)$-norm:*

$$\| h \|_{f,P_0} = \sqrt{\sup_{0 \leq \delta < \delta'} \int h^2(f(x) + \delta, x) dP_0(x)},$$

*where $\delta'$ can be chosen arbitrarily small. Let $e_{n,B_n} \equiv \{D^*_{f_{n,B_n},\epsilon_n}(Q^*_{n,B_n}, G_{n,B_n}) - D^*_{f_{n,B_n},\epsilon_n}(Q, G_0)\}$, where $Q$ is the limit of $Q_n$. If $G_n = G_0$, then $\bar{Q} \neq \bar{Q}_0$ is allowed, else $Q = Q_0$. Let $r(n) = \parallel e_{n,B_n} \parallel_{P_0}$. Let $f$ represent the limit of $\hat{f}(P_n)$, and*

$$\sigma_0^2(f) = \int_y k^2(y) dy E_{P_0} \frac{E_{P_0}((Y - \bar{Q}_0(A, X))^2 \mid A = f(X), X)}{g_0(f(X) \mid X)}.$$

*Let $\epsilon_n$ be a sequence converging to zero.*

*We make the following assumptions:*

$$\begin{aligned}
\sqrt{n\epsilon_n} E_{B_n} \parallel g_{n,B_n} - g_0 \parallel_{f_{n,B_n},P_0} \parallel \bar{Q}^*_{n,B_n} - \bar{Q} \parallel_{f_{n,B_n},P_0)} &= o_P(1) &(4)\\
\log r(n)^{-1} \max(\parallel \bar{Q}_n - \bar{Q} \parallel_{f_{n,B_n},P_0}, \parallel g_n - g_0 \parallel_{f_{n,B_n},P_0})) &= o_P(1) &(5)\\
\epsilon_n P_0 \{D^*_{f_{n,B_n},\epsilon_n}(Q, G_0)\}^2 \rightarrow \sigma_0^2(f), & &(6)
\end{aligned}$$

*as $n \rightarrow \infty$.*

*Then, $\sqrt{n\epsilon_n}(V_n^*(\hat{f}, \epsilon_n) - V_0(\hat{f}, \epsilon_n))$ converges in distribution to $N(0, \sigma_0^2(f))$, where $f$ is the limit of $\hat{f}(P_n)$. An asymptotically valid 0.95-confidence interval for $V_0(\hat{f}, \epsilon_n)$ is given by*

$$V_n^*(\hat{f}, \epsilon_n) \pm 1.96(n\epsilon_n)^{-0.5} \sigma_n(f_n),$$

*where $\sigma_n(f_n)$ is a consistent estimator of $\sigma_0(f)$.*

Regarding condition (4), even when the models for $(\bar{Q}_0, g_0)$ are nonparametric and only assume that $(\bar{Q}_0, g_0)$ are cadlag functions with finite variation norm, we can construct a highly adaptive lasso estimator $(\bar{Q}_n, g_n)$ that converges to the true $(\bar{Q}_0, g_0)$ w.r.t. $L^2(P_0)$-norm at a rate faster $n^{-1/4}$ (van der Laan, 2016). Therefore, this condition (4) is a condition that can be achieved for nonparametric models and even if $\epsilon_n$ converges very slowly to zero or is fixed. Thus, the conditions of the theorem do not necessarily put restrictions on the rate at which $\epsilon_n$ converges to zero. Regarding condition (5), for any polynomial rate $r(n)$, $\log r(n)^{-1}$ behaves as $\log n$. Thus, as long as we estimate $\bar{Q}$ and $g_0$ at a rate faster than $\log n$ w.r.t. $\parallel \cdot \parallel_{f_{n,B_n},P_0}$, condition (5) holds. Condition (6) essentially always holds.

**Proof:** For this $\epsilon$-specific one-step CV-TMLE $V_n^*(\hat{f}, \epsilon)$, we have the following identity:

$$\begin{aligned}
V_n^*(\hat{f}, \epsilon_n) - V_0(\hat{f}, \epsilon_n) &= E_{B_n}(P^1_{n,B_n} - P_0) D^*_{f_{n,B_n},\epsilon_n}(Q^*_{n,B_n}, G_{n,B_n})\\
&\quad + E_{B_n} R_{20,f_{n,B_n},\epsilon_n}(Q^*_{n,B_n}, G_{n,B_n}, Q_0, G_0).
\end{aligned}$$

We will prove

$$\begin{aligned}\sqrt{n\epsilon_n}E_{B_n}(P^1_{n,B_n}-P_0)e_{n,B_n} &= o_P(1)\\ \sqrt{n\epsilon_n}E_{B_n}R_{20,f_{n,B_n},\epsilon_n}(Q^*_{n,B_n},G_{n,B_n},Q_0,G_0) &= o_P(1).\end{aligned}$$

In the special case that $G_n=G_0$ is known we have $R_{20,f,\epsilon}(Q^*_n,G_n,Q_0,G_0)=0$, so that we only need to show the first equality at a possibly misspecified $\bar{Q}$. Regarding the first condition we apply Lemma 12 in van der Laan (HAL), which proves

$$\sqrt{n}E_{B_n}(P^1_{n,B_n}-P_0)e_{n,B_n}=O_P(r(n)(1-\log r(n))).$$

We have that $\parallel e_{n,B_n}\parallel_{P_0}$ behaves as $\epsilon_n^{-0.5}$ times the maximum of $\parallel \bar{Q}^*_{n,B_n}-\bar{Q}_0\parallel_{f_{n,B_n},P_0}$ and $\parallel g_{n,B_n}-g_0\parallel_{f_{n,B_n},P_0}$. By condition (5) this proves that $\sqrt{n\epsilon_n}E_{B_n}(P^1_{n,B_n}-P_0)e_{n,B_n}$ converges to zero in probability.

Thus, we now have

$$(n\epsilon_n)^{0.5}\left(V^*_n(\hat{f},\epsilon_n)-V_0(\hat{f},\epsilon_n)\right)=(n\epsilon_n)^{0.5}E_{B_n}(P^1_{n,B_n}-P_0)D^*_{f_{n,B_n},\epsilon_n}(Q_0,G_0)+o_P(1).$$

For each $B_n$, $(P_{n,B_n}-P_0)D^*_{f_{n,B_n},\epsilon_n}(Q_0,G_0)$ equals a sum of $n$ i.i.d. mean zero random variables. Therefore, if for each $B_n$, condition (6) holds, then

$$\sqrt{n\epsilon_n}(V^*_n(\hat{f},\epsilon_n)-V_0(\hat{f},\epsilon_n)\Rightarrow_d N(0,\sigma^2_0(f),$$

where $f$ is the limit of $f_{n,B_n}$. This completes the proof. □

### 3.4 Asymptotic normality of CV-TMLE as an estimator of risk of candidate estimator

We have the following result regarding the approximation $V_0(\hat{f},\epsilon)-V_0(\hat{f})$.

**Lemma 1** *Let* $V_0(\hat{f})=E_{B_n}V_0(f_{n,B_n})$ *and* $V_0(\hat{f},\epsilon)=E_{B_n}E_0\int_a k_{f_{n,B_n}(X),\epsilon}(a)\bar{Q}_0(a,X)da$. *Let*

$$B_0(f)=B_0(f,J)\equiv\int k(y)y^J dy E_{B_n}E_0\frac{1}{J!}\bar{Q}_0^{(J)}(f(X),X).$$

*Suppose* $P_0\in\mathcal{M}(J)$, $\epsilon_n\rightarrow 0$, *and* $E_{B_n}B_0(f_{n,B_n})\rightarrow B_0(f,J)$ *for a limit rule* $f$.

*Then,*

$$\epsilon^{-J}(V_0(\hat{f},\epsilon)-V_0(\hat{f}))\rightarrow B_0(f,J)\ \textit{as}\ \epsilon\rightarrow 0.$$

**Proof:** Using a $J$-th order Tailor expansion around $f(X)$, we obtain

$$\begin{aligned}
V_0(f,\epsilon)-V_0(f) &= E_0\int_a k_{f(X),\epsilon}(a)(\bar{Q}_0(a,X)-\bar{Q}_0(f(X),X))da \\
&= E_0\int_a dak_{f(X),\epsilon}(a)\sum_{j=1}^{J-1}\frac{1}{j!}\bar{Q}_0^{(j)}(f(X),X)(a-f(X))^j \\
&\quad +E_0\int_a dak_{f(X),\epsilon}(a)(a-f(X))^J\frac{1}{J!}\bar{Q}_0^{(J)}(\xi(a,X),X)+o(\epsilon^J) \\
&= \sum_{j=1}^{J-1}\frac{\epsilon^j}{j!}E_0\bar{Q}_0^{(j)}(f(X),X)\int k(y)y^j dy \\
&\quad +\epsilon^J E_0\int dak_{f(X),\epsilon}(a)(a-f(X))^J\frac{1}{J!}\bar{Q}_0^{(J)}(\xi(a,X),X)+o(\epsilon^J) \\
&= \epsilon^J E_0\int dak_{f(X),\epsilon}(a)(a-f(X))^J\frac{1}{J!}\bar{Q}_0^{(J)}(\xi(a,X),X)+o(\epsilon^J) \\
&= \epsilon^J E_0\int k(y)y^J dy\frac{1}{J!}\bar{Q}_0^{(J)}(f(X),X)+o(\epsilon^J) \\
&\equiv \epsilon^J B_0(f,J)+o(\epsilon^J),
\end{aligned}$$

where $\xi(a,X)$ is a value between $a$ and $f(X)$. So we can conclude that $V_P(f,\epsilon)-V_P(f)=O(\epsilon^J)$ and specifically that $\epsilon^{-J}(V_P(f,\epsilon)-V_P(f))\rightarrow B_0(f,J)$. In the same manner, it follows that

$$V_0(\hat{f},\epsilon)-V_0(\hat{f})=\epsilon^J E_{B_n}E_0B_0(f_{n,B_n},J)+o(\epsilon^J),$$

and thus that $\epsilon^{-J}(V_0(\hat{f},\epsilon)-V_0(\hat{f}))\rightarrow B_0(f,J)$. This completes the proof. □

Thus, for the sake of optimizing the rate of convergence of $V_n^*(\hat{f},\epsilon_n)$ as an estimator of $V_0(\hat{f})$, we should balance the square-bias $\epsilon^{2J}$ with the variance $1/(n\epsilon)$ of $V_n^*(\hat{f},\epsilon)$, resulting in the optimal bandwidth $\epsilon_n=n^{-1/(2J+1)}$ and corresponding optimal rate of convergence $n^{-(2J)/(2J+1)}$. This proves the following theorem, which is a corollary of Theorem 1 and the above lemma.

**Theorem 2** *Assume $P_0\in\mathcal{M}(J)$, (4), (5), (6), and $E_{B_n}B_0(f_{n,B_n})\rightarrow_p B_0(f,J)$ for a limit rule $f$. Let $\epsilon_n=n^{-1/(2J+1)}$. Then,*

$$n^{-J/(2J+1)}(V_n^*(\hat{f},\epsilon_n)-V_0(\hat{f}))\Rightarrow_d N(B_0(f,J),\sigma_0(f)),$$

*where $f$ is the limit of $\hat{f}(P_n)$.*

*If $\epsilon_n$ converges faster to zero than this rate so that $\epsilon_n/n^{-1/(2J+1)}\rightarrow 0$, then*

$$(n\epsilon_n)^{0.5}(V_n^*(\hat{f},\epsilon_n)-V_0(\hat{f}))\Rightarrow_d N(0,\sigma_0(f)).$$

*If $P_0 \in \mathcal{M}(J_0)$ for a $J_0 \leq J$, then we have that for $\epsilon_n = n^{-1/(2J_0+1)}$*

$$n^{-J_0/(2J_0+1)}(V_n^*(\hat{f}, \epsilon_n) - V_0(\hat{f})) \Rightarrow_d N(B_0(f, J_0), \sigma_0(f)).$$

The latter result in terms of the unknown smoothness $J_0$ is nicer since it does not rely on having to assume that $P_0 \in \mathcal{M}(J)$. However, in this case the optimal rate for the bandwidth $\epsilon$ is unknown. Fortunately, our data adaptive selector of $\epsilon$ automatically finds this optimal rate as proved below.

# 4 A data adaptive selector of $\epsilon$.

In the first subsection we define the selector, and in the second subsection we show that the selector achieves the optimal rate corresponding with the unknown underlying smoothness level $J_0$ of $\bar{Q}_0$.

## 4.1 Minimizing right upper bound of confidence interval for true $\epsilon$-risk of candidate estimator

Let $\sigma_n^2$ be an estimator of $\sigma_0^2(f) = \int_y k^2(y)dyE_{P_0}\frac{E_{P_0}((Y-\bar{Q}_0(A,X))^2|A=f(X),X)}{g_0(f(X)|X)}$. For example, if $Y$ is binary, we can use the following cross-validated estimator:

$$\int k^2(y)dyE_{B_n}E_{P^1_{n,B_n}}\frac{\bar{Q}_{n,B_n}(1-\bar{Q}_{n,B_n})(f_{n,B_n}(X),X)}{g_{n,B_n}(f_{n,B_n}(X)\mid X)}.$$

Alternatively, one can use

$$\sigma_n^2 = \epsilon_n' E_{B_n}P^1_{n,B_n}D^*_{f_{n,B_n},\epsilon_n'}(\bar{Q}_{n,B_n}, G_{n,B_n})^2$$

for a choice $\epsilon_n'$. Our proposed selector $\epsilon_n$ is defined by

$$\epsilon_n = \arg\min_{\epsilon}\left\{V_n^*(\hat{f},\epsilon) + Z(1-\alpha/2)\frac{\sigma_n}{(n\epsilon)^{0.5}}\right\}. \tag{7}$$

In other words, $\epsilon_n$ is defined as the minimizer of the right-upper bound of the $1-\alpha/2$-confidence interval $V_n^*(\hat{f},\epsilon) \pm Z(1-\alpha/2)\sigma_n/(n\epsilon)^{0.5}$ for $V_0(\hat{f},\epsilon)$.

## 4.2 Asymptotic theory

For the purpose of simplifying our analysis, we focus on the above selector for the case that $V_n^*(\hat{f},\epsilon)$ is given by the double robust efficient (for fixed $\epsilon$) CV-AIPTW estimator:

$$V_n(\hat{f},\epsilon) = E_{B_n}P^1_{n,B_n}D_{f_{n,B_n},\epsilon}(Q_{n,B_n}, G_{n,B_n}), \tag{8}$$

where

$$D_{f,\epsilon}(Q,G)(O) = \frac{k_{f(X),\epsilon}(A)}{g(A \mid X)}(Y - \bar{Q}(A,X)) + \int_a k_{f(X),\epsilon}(a)\bar{Q}(a,X)da.$$

In other words, $D_{f,\epsilon}$ is the non-centered efficient influence curve $D^*_{f,\epsilon} = D_{f,\epsilon} - V(f,\epsilon)$. This estimator $V_n(\hat{f},\epsilon)$ has the same asymptotic behavior as the CV-TMLE. In fact, the CV-TMLE can be represented by (8) but with $Q_{n,B_n}$ replaced by the targeted $Q^*_{n,B_n}$. Since $Q_{n,B_n}$ is not targeted it does depend on $\epsilon$, while the targeted $\bar{Q}^*_{n,B_n}$ in the CV-TMLE does also depend on the clever covariate and thereby is a function of $\epsilon$. As a consequence, our derivative w.r.t. $\epsilon$ of the criterion minimized by $\epsilon_n$ would also include the derivative w.r.t the clever covariate, creating additional terms we chose to avoid, We expect that the same theorem applies to the selector using the CV-TMLE, but that remains to be investigated.

Our theorem below only relies on the derivative of the criterion (i.e., right-upper bound of confidence interval) being zero at $\epsilon_n$, which, of course, holds for our $\epsilon_n$ whenever it is an interior minimum. As a consequence, our theorem immediately generalizes to any solution $\epsilon_n$ solving:

$$\frac{d}{d\epsilon_n} V_n(\hat{f},\epsilon_n) = C\frac{d}{d\epsilon_n}\left\{\sigma_n/(n\epsilon_n)^{0.5}\right\}$$

for some constant $C > 0$. In other words, wen can prove the same theorem for an $\epsilon_n$ for which the infinitesimal change in the estimator $V_n(\hat{f},\epsilon)$ is proportional to the infinitesimal change in the standard error of the estimator.

**Theorem 3** *Consider the definition (7) of $\epsilon_n$ with $V^*_n(\hat{f},\epsilon)$ being the CV-AIPTW estimator (8). We define*

$$\begin{aligned}
D^*_{1,f,\epsilon}(Q,G) &= \frac{k_{1,f(X),\epsilon}(A)}{g(A \mid X)}(Y - \bar{Q}(A,X)) + \int k_{1,f(X),\epsilon}(a)\bar{Q}(a,X)da \\
k_{1,f(X),\epsilon}(a) &= \epsilon\frac{d}{d\epsilon}\left(\epsilon^{-1}k((a - f(X))/\epsilon)\right) \\
&= -\epsilon^{-1}k_{f(X),\epsilon}(a) + \epsilon^{-2}k^{(1)}((a - f(X))/\epsilon)(a - f(X)) \\
e_{1,n,B_n} &\equiv \{D^*_{1,f_{n,B_n},\epsilon_n}(Q_{n,B_n}, G_{n,B_n}) - D^*_{1,f_{n,B_n},\epsilon_n}(Q,G_0)\} \\
r_1(n) &= \parallel e_{1,n,B_n} \parallel_{P_0} \\
Z_n &= \sqrt{n}(P_n - P_0)\epsilon^{0.5}D^*_{1,f,\epsilon}(Q,G_0) \\
\sigma^2_{10}(f) &\equiv \lim_{\epsilon\to 0} VAR\left(\epsilon^{0.5}D^*_{1,f,\epsilon}(Q,G_0)\right),
\end{aligned}$$

*where $Q = (Q_{X,0}, \bar{Q})$ is the limit of $Q_n$. If $G_0$ is known and $G_n = G_0$, then $\bar{Q} \neq \bar{Q}_0$ is allowed, else $Q = Q_0$. Let $f$ represent the limit of $\hat{f}(P_n)$,*
**Assumptions:** *Let $P_0 \in \mathcal{M}(J_0)$, and let $\epsilon_n \rightarrow 0$ be a solution of*

$$\frac{d}{d\epsilon_n} V_n(\hat{f}, \epsilon_n) = 0.5Z(1-\alpha/2)\sigma_n n^{-0.5} \epsilon_n^{-3/2}.$$

*In addition, assume*

$$\begin{aligned}
E_{B_n} \parallel g - g_0 \parallel_{P_0, f_{n,B_n}} \parallel \bar{Q}^*_{n,B_n} - \bar{Q}_0 \parallel_{P_0, f_{n,B_n}} &= o_P(n^{-1/2}) && (9)\\
\log r_1(n)^{-1} \max(\parallel \bar{Q}_n - \bar{Q} \parallel_{f_{n,B_n}, P_0}, \parallel g_n - g_0 \parallel_{f_{n,B_n}, P_0})) &= o_P(1) && (10)\\
E_0 \bar{Q}_0^{(J_0)}(f_{n,B_n}(X), X) - E_0 \bar{Q}_0^{(J_0)}(f(X), X) &\rightarrow_p 0 && (11)\\
\epsilon_n P_0 \{ D^*_{1, f_{n,B_n}, \epsilon_n}(Q, G_0) \}^2 &\rightarrow_p \sigma^2_{10}(f). && (12)
\end{aligned}$$

*Then,*

$$n \epsilon_n^{2J_0+1} J_0^2 B_0(f, J_0)^2 = \{ Z_n - 0.5Z(1-\alpha/2)\sigma_n \}^2 + o_P(1).$$

**Proof:** We have

$$\frac{d}{d\epsilon_n} V_n(\hat{f}, \epsilon_n) = 0.5Z(1-\alpha/2)\sigma_n n^{-0.5} \epsilon_n^{-3/2}.$$

We also have

$$\begin{aligned}
V_n(\hat{f}, \epsilon_n) - V_0(\hat{f}, \epsilon) &= E_{B_n}(P^1_{n,B_n} - P_0) D^*_{f_{n,B_n}, \epsilon_n}(Q_{n,B_n}, G_{n,B_n}) \\
&\quad + R_{20, f_{n,B_n}, \epsilon_n}(Q_{n,B_n}, G_{n,B_n}.Q_0, G_0).
\end{aligned}$$

Thus,

$$\begin{aligned}
-\frac{d}{d\epsilon_n} V_0(\hat{f}, \epsilon_n) &= \epsilon_n^{-1} E_{B_n}(P^1_{n,B_n} - P_0) D^*_{1, f_{n,B_n}, \epsilon_n}(Q_{n,B_n}, G_{n,B_n}) \\
&\quad + \epsilon_n^{-1} R_{20,1, f_{n,B_n}, \epsilon_n}(Q_{n,B_n}, G_{n,B_n}, Q_0, G_0) \\
&\quad - 0.5Z(1-\alpha/2)\sigma_n n^{-0.5} \epsilon_n^{-3/2},
\end{aligned}$$

where

$$R_{20,1,f,\epsilon}(Q, G, Q_0, G_0) = E_{P_0} \int k_{1, f(X), \epsilon}((a) \frac{(g - g_0)}{g g_0}(a \mid X)(\bar{Q} - \bar{Q}_0)(a, X) da.$$

Note that $k_{1,f(X),\epsilon}$ is bounded in absolute value by $O(\epsilon^{-1})$, since $\mid (a - f(X)) \mid \leq \epsilon$ when $k^{(1)}((a - f(X))/\epsilon) \neq 0$. As a consequence, $\mathrm{VAR} D^*_{1,f,\epsilon}(Q_0, G_0) = O(\epsilon^{-1})$.

We can analyze the empirical process term in exactly the same manner as we analyzed the empirical process term $E_{B_n}(P^1_{n,B_n} - P_0)D^*_{f_{n,B_n},\epsilon}(Q^*_{n,B_n}, G_{n,B_n})$ in the CV-TMLE (Theorem 1). This proves Lemma 2. By application of Lemma 2 we have

$$\sqrt{n\epsilon}E_{B_n}(P^1_{n,B_n} - P_0)D^*_{1,f_{n,B_n},\epsilon}(Q_{n,B_n}, G_{n,B_n}) = Z_n + o_P(1).$$

Under the assumption that $P_0 \in \mathcal{M}(J_0)$, application of Lemma 3 proves that

$$\frac{d}{d\epsilon_n}V_0(\hat{f}, \epsilon_n) = \epsilon_n^{J_0-1}J_0B_0(f, J_0) + o_P(\epsilon_n^{J_0}).$$

So we have shown

$$\begin{aligned} -\epsilon_n^{J_0-1}J_0B_0(f, J_0) + o_P(\epsilon_n^{J_0}) &= \epsilon_n^{-1}(n\epsilon_n)^{-0.5}(Z_n + o_P(1)) \\ &\quad +\epsilon_n^{-1}E_{B_n}R_{20,1,f_{n,B_n},\epsilon_n}(Q_{n,B_n}, G_{n,B_n}, Q_0, G_0) \\ &\quad -0.5Z(1-\alpha/2)\sigma_n n^{-0.5}\epsilon_n^{-3/2}, \end{aligned}$$

where $Z_n \Rightarrow N(0, \sigma^2_{01}(f))$. Multiplying both sides with $\epsilon_n$ yields:

$$\begin{aligned} -\epsilon_n^{J_0}J_0B_0(f, J_0) + o_P(\epsilon_n^{J_0+1}) &= (n\epsilon_n)^{-0.5}(Z_n + o_P(1)) \\ &\quad +E_{B_n}R_{20,1,f_{n,B_n},\epsilon_n}(Q_{n,B_n}, G_{n,B_n}, Q_0, G_0) \\ &\quad -0.5Z(1-\alpha/2)\sigma_n(n\epsilon)^{-0.5}. \end{aligned}$$

We now have to analyze $R_{20,1,f_{n,B_n},\epsilon_n}$. Note that $k_{1,f(X),\epsilon}(a) = -k_{I,f(X),\epsilon}(a) + k_{II,f(X),\epsilon}(a)$ is a sum of two functions. Consider the two resulting integrals and carry out the substitution $(a - f_{n,B_n}(X))/\epsilon = y$. The first term can be bounded as follows:

$$\begin{aligned} &\mid E_{P_0}\int k(y)\tfrac{(g-g_0)}{gg_0}(f(X)+\epsilon y \mid X)(\bar{Q}-\bar{Q}_0)(f(X)+\epsilon y, X)dy \mid \\ &\leq C\int_{x,y} \mid k(y) \mid\mid (g-g_0) \mid (f(x)+\epsilon y \mid x) \mid \bar{Q}-\bar{Q}_0 \mid (f(x)+\epsilon y, x)dP_0(x)dy \\ &\leq C\sqrt{\int_{x,y} \mid k(y) \mid (g-g_0)^2(f(x)+\epsilon y \mid x)dP_0(x)dy}\times \\ &\sqrt{\int_{x,y} \mid k(y) \mid\mid \bar{Q}-\bar{Q}_0 \mid^2 (f(x)+\epsilon y, x)dP_0(x)dy} \\ &\leq C \parallel g-g_0 \parallel_{P_0,f}\parallel \bar{Q}-\bar{Q}_0 \parallel_{P_0,f}, \end{aligned}$$

for some $C < \infty$ depending on the kernel $k$ and the lower bound $\delta$ for $g_0$. Similarly, the second term can be bounded as follows:

$$\begin{aligned} &\mid E_{P_0}\int k^{(1)}(y)y\tfrac{(g-g_0)}{gg_0}(f(X)+\epsilon y \mid X)(\bar{Q}-\bar{Q}_0)(f(X)+\epsilon y, X)dy \mid \\ &\leq C\int_{x,y} \mid k^{(1)}(y)y \mid\mid (g-g_0) \mid (f(x)+\epsilon y \mid x) \mid \bar{Q}-\bar{Q}_0 \mid (f(x)+\epsilon y, x)dP_0(x)dy \\ &\leq C\sqrt{\int_{x,y} \mid k^{(1)}(y)y \mid (g-g_0)^2(f(x)+\epsilon y \mid x)dP_0(x)dy} \\ &\sqrt{\int_{x,y} \mid k(y) \mid\mid \bar{Q}-\bar{Q}_0 \mid^2 (f(x)+\epsilon y, x)dP_0(x)dy} \\ &\leq C \parallel g-g_0 \parallel_{P_0,f}\parallel \bar{Q}-\bar{Q}_0 \parallel_{P_0,f}. \end{aligned}$$

Thus, this proves

$$E_{B_n} R_{20,1,f_{n,B_n},\epsilon_n}(Q_{n,B_n}, G_{n,B_n}, Q_0, G_0) \leq C E_{B_n} \parallel g - g_0 \parallel_{P_0,f_{n,B_n}} \parallel \bar{Q}^*_{n,B_n} - \bar{Q}_0 \parallel_{P_0,f_{n,B_n}} .$$

By condition (13) the last term is $o_P(n^{-1/2})$. We have now shown:

$$-\epsilon_n^{J_0} J_0 B_0(f, J_0) + o_P(\epsilon_n^{J_0+1}) = (n\epsilon_n)^{-0.5} \left\{ Z_n - 0.5 Z(1-\alpha/2)\sigma_n + o_P(1) \right\} + o_P(n^{-1/2}).$$

This proves that in a small enough neighborhood of 0, the solution is driven by the equation

$$-\epsilon_n^{J_0} J_0 B_0(f, J_0) = (n\epsilon_n)^{-0.5} \left\{ Z_n - 0.5 Z(1-\alpha/2)\sigma_n \right\},$$

and thus

$$n\epsilon_n^{2J_0+1} J_0^2 B_0(f, J_0)^2 = \left\{ Z_n - 0.5 Z(1-\alpha/2)\sigma_n \right\}^2 + o_P(1),$$

which completes the proof. □

Our proof of Theorem 3 relied on the following two lemmas.

**Lemma 2** *Let* $e_{1,n,B_n} \equiv \{D^*_{1,f_{n,B_n},\epsilon_n}(Q_{n,B_n}, G_{n,B_n}) - D^*_{1,f_{n,B_n},\epsilon_n}(Q, G_0)\}$, *where* $Q$ *is the limit of* $Q_n$. *If* $G_n = G_0$, *then* $\bar{Q} \neq \bar{Q}_0$ *is allowed, else* $Q = Q_0$. *Let* $r_1(n) = \parallel e_{1,n,B_n} \parallel_{P_0}$. *Let* $f$ *represent the limit of* $\hat{f}(P_n)$, *and*

$$\sigma^2_{10}(f) = \lim_{\epsilon \to 0} \epsilon^{0.5} D^*_{1,f,\epsilon}(Q_0, G_0).$$

*Let* $\epsilon_n$ *be a sequence converging to zero. We make the following assumptions:*

$$\begin{aligned} \log r_1(n)^{-1} \max(\parallel \bar{Q}_n - \bar{Q} \parallel_{f_{n,B_n},P_0}, \parallel g_n - g_0 \parallel_{f_{n,B_n},P_0})) &= o_P(1) \quad (13) \\ \epsilon_n P_0 \{D^*_{1,f_{n,B_n},\epsilon_n}(Q, G_0)\}^2 &\to \sigma^2_{10}(f), \quad (14) \end{aligned}$$

*as* $n \to \infty$. *Then,*

$$\sqrt{n\epsilon} E_{B_n}(P^1_{n,B_n} - P_0) D^*_{1,f_{n,B_n},\epsilon}(Q_{n,B_n}, G_{n,B_n}) = \sqrt{n}(P_n - P_0)\epsilon^{0.5} D^*_{1,f,\epsilon}(Q_0, G_0) + o_P(1),$$

*and the right-hand side converges in distribution to* $N(0, \sigma^2_{01}(f))$.

**Lemma 3** *Suppose that* $P_0 \in \mathcal{M}(J_0)$, *and* $E_0 \bar{Q}_0^{(J_0)}(f_{n,B_n}(X), X) \to_p E_0 \bar{Q}_0^{(J_0)}(f(X), X)$.

*Then,*

$$\frac{d}{d\epsilon} V_0(\hat{f}, \epsilon) = \epsilon^{J_0-1} J_0 B_0(f, J_0) + o_P(\epsilon^{J_0}).$$

**Proof:** We have

$$
\begin{aligned}
\frac{d}{d\epsilon}V_0(\hat{f},\epsilon) &= E_0 \int_y yk(y)\bar{Q}_0^{(1)}(f_{n,B_n}(X)+\epsilon y, X) \\
&= E_0 \left\{ \int_y yk(y)\bar{Q}_0^{(1)}(f_{n,B_n}(X),X) + \sum_{j=2}^{J_0-1} \frac{(\epsilon y)^{j-1}}{(j-1)!}\bar{Q}_0^{(j)}(f_{n,B_n}(X),X) \right\} \\
&\quad + E_0 \int yk(y)\frac{(\epsilon y)^{J_0-1}}{(J_0-1)!}\bar{Q}_0^{(J_0)}(\xi(f_{n,B_n}(X), f_{n,B_n}(X)+\epsilon y), X) \\
&= \epsilon^{J_0-1} E_0 \int yk(y)\frac{1}{(J_0-1)!}\bar{Q}_0^{(J_0)}(\xi(f_{n,B_n}(X), f_{n,B_n}(X)+\epsilon y), X) \\
&= \epsilon^{J_0-1} J_0 E_0 \int yk(y)dy \frac{1}{J_0!}\bar{Q}_0^{(J_0)}(f_{n,B_n}(X),X) + o_P(\epsilon^{J_0}) \\
&= \epsilon^{J_0-1} J_0 E_0 \int yk(y)dy \frac{1}{J_0!}\bar{Q}_0^{(J_0)}(f(X),X) + o_P(\epsilon^{J_0}) \\
&= \epsilon^{J_0-1} J_0 B_0(f,J_0) + o_P(\epsilon^{J_0}).
\end{aligned}
$$

This completes the proof. □

**Remark** Thus, for this selector $\epsilon_n$, we have that $n^{-J_0/(2J_0+1)}(V_n^*(\hat{f},\epsilon_n) - V_0(\hat{f}))$ converges to a normal distribution with variance $\sigma_0^2(f)$ and a bias term involving $B_0(f,J_0)$ depending on the $J_0$-th derivative of $\bar{Q}_0$. For the purpose of constructing a confidence interval for $V_0(\hat{f})$, one might want to under-smooth so that this bias term in the normal limit distribution disappears. For example, one might use $\epsilon_n/\log n$ as data adaptive selector of $\epsilon$ in $V_n^*(\hat{f},\epsilon)$ for the purpose of constructing a confidence interval for $V_0(\hat{f})$, while we still use $\epsilon_n$ to get our best estimator of $V_0(\hat{f})$ in the cross-validation selector of the super-learner.

# 5 Super-learner of optimal rule

## 5.1 Creating a large library of candidate estimators

Any candidate estimator $\bar{Q}_n$ of $\bar{Q}_0$ implies also a corresponding estimator $f_n(X) = \arg\max_a \bar{Q}_n(a,X)$. For our purpose the better estimators would be the ones that do a good job in estimation of the conditional treatment effect $\bar{Q}_0(a,X) - \bar{Q}_0(0,X)$ relative to some baseline treatment 0. For example, we could use the TMLE of a parametric model $m_\beta(a,x)$ for $\bar{Q}_0(a,X) - \bar{Q}_0(0,X)$ as presented in the literature (Yu, Wang, vdL papers). These models are often called semi-parametric regression models in which only the treatment effect component is modeled, while $\bar{Q}_0(0, X)$ is left unspecified. Each different choice of parametric form implies now a candidate estimator of $f_0$. One could also construct the following type of two-stage estimator. First compute an estimator of $E(Y \mid A = 0, X)$, for example, by using a super-learner regressing $Y$ on $X$ using a kernel-weighted loss function $h^{-1}k(A/h)(Y - \bar{Q}(X, 0))^2$ with bandwidth $h$, but other loss functions such as $EY_{f(Q)}$ could be considered as well, where $f(Q)$ is the rule implied by $Q$. Subsequently, we use this estimator as off-set and run a regression of the residual $Y - \bar{Q}_n^0(0, X)$ on $(A, X)$ fitting a model $Am(X)$ where $m$ can be arbitrary. These are just examples. Due to all the different choices in any of these estimation procedure, it is not hard to generate a large and diverse collection of candidate estimators.

## 5.2 CV-TMLE cross-validation selector

Having defined a cross-validated TMLE of $V_0(\hat{f}, \epsilon)$ for any given candidate estimator $\hat{f}$ of $f_0$, we are now ready to define a super-learner. Consider a collection of candidate estimators $\hat{f}_m$, $m = 1, \ldots, M_n$, of $f_0$. Let

$$m_n \equiv \arg\min_m V_n^*(\hat{f}_m, \epsilon_n)$$

be the cross-validation selector of $m$ that minimizes the above defined CV-TMLE of $V_0(\hat{f}, \epsilon_n)$. The discrete super-learner is defined as $f_n \equiv \hat{f}_{m_n}(P_n)$.

Consider now a family $\{\hat{f}_\alpha : \alpha\}$ of candidate estimators, where $\hat{f}_\alpha$ is a candidate estimator implied by $\hat{f}_m$, $m = 1, \ldots, M_n$, and a vector $\alpha$. For example, one might define $\hat{f}_\alpha = \frac{1}{M} \sum_{m=1}^{M_n} \hat{f}_m$ as the average of the treatment selected by the different candidate rules $\hat{f}_m$. We can then define the cross-validation selector of $\alpha$ as

$$\alpha_n \equiv \arg\min_\alpha V_n^*(\hat{f}_\alpha, \epsilon_n).$$

This now defines the more aggressive super-learner $f_n \equiv \hat{f}_{\alpha_n}(P_n)$.

We focus on the analysis of the discrete super-learner, which also implies results for the continuous super-learner by discretizing the $\alpha$-space

## 5.3 Defining CV-TMLE cross-validation selector as loss-function based cross-validation selector

We note that the CV-TMLE of $V_0(\hat{f}, \epsilon)$ solves the cross-validated efficient influence curve equation $E_{B_n} P^1_{n,B_n} D^*_{f_{n,B_n},\epsilon}(Q^*_{n,B_n}, G_{n,B_n}) = 0$. This shows that we can represent our CV-TMLE of $V_0(\hat{f},\epsilon)$ as a cross-validated empirical risk of a loss function:

$$V_n^*(\hat{f},\epsilon) = E_{B_n} P_{n,B_n}^1 L_{\bar{Q}_{n,B_n}^*,G_{n,B_n},\epsilon}(\hat{f}(P_{n,B_n}^0)),$$

where

$$L_{\bar{Q},G,\epsilon}(f) = \frac{k_{f(X),\epsilon}(A)}{g(A \mid X)}(Y - \bar{Q}(A,X)) + \int k_{f(X),\epsilon}(a)\bar{Q}(a,X)da.$$

Here $L_{\bar{Q},G,\epsilon}(f)$ can be viewed as a loss function indexed by nuisance parameters $(\bar{Q},G)$.

Even though this is not the manner a CV-TMLE is computed, this representation will come handy when establishing an oracle inequality for the CV-TMLE cross-validation selector, since it allows us to represent the cross-validation selector as one based on minimizing the cross-validated empirical risk of a loss function that is indexed by a nuisance parameter. Such cross-validation selectors have been studied in detail in van der Laan, Dudoit (2003) and subsequent articles.

Thus, the CV-TMLE cross-validation selector $m_n$ can be represented as:

$$m_n = \arg\min_m E_{B_n} P_{n,B_n}^1 L_{\bar{Q}_{n,B_n}^*,G_{n,B_n},\epsilon_n}(\hat{f}(P_{n,B_n}^0)).$$

The loss function is double robust in the sense that

$$V_0(\hat{f},\epsilon) = E_{B_n} P_0 L_{\bar{Q},G,\epsilon}(\hat{f}(P_{n,B_n}^0)),$$

if either $\bar{Q} = \bar{Q}_0$ or $G = G_0$. That is, $L_{\bar{Q},G,\epsilon}$ is a loss function for $f_{0,\epsilon} = \arg\min_f P_0 L_{\bar{Q},G,\epsilon}(f)$ when one of nuisance parameters $(\bar{Q},G)$ is correctly specified.

## 5.4 The CV-IPTW cross-validation selector

If we select $\bar{Q} = 0$, then we obtain

$$V_0(f,\epsilon) = E_0 L_{G_0,\epsilon}(f),$$

where

$$L_{G_0,\epsilon}(f) \equiv \frac{k_{f(X),\epsilon}(A)}{g_0(A \mid X)} Y$$

is the IPTW-loss function. The corresponding cross-validated empirical risk of $V_0(\hat{f},\epsilon)$ is then given by

$$V_{n,IPTW}(\hat{f},\epsilon) = E_{B_n} P_{n,B_n}^1 L_{G_{n,B_n},\epsilon}(\hat{f}(P_{n,B_n}^0)).$$

In that case, the corresponding cross-validation selector is thus given by:

$$m_{n,IPTW} = \arg\min_m E_{B_n} P^1_{n,B_n} L_{G_{n,B_n},\epsilon}(\hat{f}_m(P^0_{n,B_n}).$$

The CV-TMLE estimator $V_n^*(\hat{f}, \epsilon)$ of $V_0(\hat{f}, \epsilon)$ is generally a more efficient estimator than this IPTW-estimator $V_{n,IPTW}(\hat{f}, \epsilon)$, and, since it is a substitution estimator, it is guaranteed to respect the global constraints (such as constraint on possible values of the risk function). Therefore, the performance of the cross-validation selector $m_n$ based on the CV-TMLE can be expected to generally perform better than $m_{n,IPTW}$ in finite samples.

### 5.5 The CV-AIPTW cross-validation selector

If we replace the TMLE $\bar{Q}^*_{n,B_n}$ by the non-targeted initial estimator $\bar{Q}_{n,B_n}$ in the CV-TMLE selector, then we obtain the CV-AIPTW selector:

$$m_{n,AIPTW} = \arg\min_m E_{B_n} P^1_{n,B_n} L_{\bar{Q}_{n,B_n},G_{n,B_n},\epsilon_n}(\hat{f}(P^0_{n,B_n}).$$

Note that the CV-AIPTW estimator of risk only differs from the CV-TMLE by using the initial estimator $\bar{Q}_{n,B_n}$ instead of the targeted $\bar{Q}^*_{n,B_n}$ (thereby not guaranteeing that it is a substitution estimator).

## 6 Oracle inequality for super-learner

### 6.1 Risk-based dissimilarity

Let $f_{0,\epsilon} = \arg\min_f V_0(f, \epsilon)$. We define the $\epsilon$-risk-based dissimilarity as:

$$d_{0,\epsilon}(f, f_{0,\epsilon}) = V_0(f, \epsilon) - V_0(f_{0,\epsilon}, \epsilon) \geq 0.$$

We also define the risk based dissimilarity

$$d_0(f, f_0) = V_0(f) - V_0(f_0) \geq 0.$$

If $f_0(X)$ is an interior minimum as stated in (2) and $P_0 \in \mathcal{M}(J = 2)$, then a second order Tailor expansion proves that $d_0(f, f_0) = O(\int (f - f_0)^2 dP_0(x))$. The oracle inequalities in this section are not of interest in this special case due to the remainder term not being good enough. We refer to the Appendix for a presentation of the oracle inequalities for the case that the risk dissimilarity behaves as a square of a norm.

## 6.2 $\epsilon$-oracle inequality

We have the following theorem (e.g., van der Laan, Dudoit (2003), van der Vaart et al. (2006))

**Theorem 4** *Let $P_0 \in \mathcal{M}(J=0)$. Assume $M_1 = \sup_{P \in \mathcal{M}(0)} \sup_{f \in \mathcal{F}} \sup_o \mid L_{\bar{Q},G}(f)(o) \mid < \infty$. Assume that $G_0$ is known and $G_n = G_0$. Let $m_n$ be either the CV-TMLE, CV-AIPTW, CV-IPTW selector. Then,*

$$\begin{aligned} Ed_{0,\epsilon}(\hat{f}_{m_n}(P^0_{n,B_n}), f_{0,\epsilon}) &\leq E \min_m E_{B_n} d_{0,\epsilon}(\hat{f}_m(P^0_{n,B_n}), f_{0,\epsilon}) \\ &\quad + C(M_1) \frac{\sqrt{\log M_n + \log n}}{(n\epsilon p)^{0.5}}. \end{aligned}$$

*In the case that $m_n$ equals $m_{n,IPTW}$ or $m_{n,AIPTW}$ one can remove the $\log n$ in the remainder term.*

## 6.3 Approximation error of $\epsilon$-dissimilarity

**Theorem 5** *Let $P_0 \in \mathcal{M}^J$. Then,*

$$d_{0,\epsilon}(\hat{f}, f_{0,\epsilon}) - d_0(\hat{f}, f_0) = O(\epsilon^J).$$

**Proof:** It suffices to prove that $d_{0,\epsilon}(f, f_{0,\epsilon}) - d_0(f, f_0) = O(\epsilon^J)$. Firstly, note that

$$\begin{aligned} d_{0,\epsilon}(f, f_{0,\epsilon}) - d_0(f, g_0) &= (V_0(f,\epsilon) - V_0(f)) + (V_0(f_{0,\epsilon},\epsilon) - V_0(f_0,\epsilon)) \\ &\quad + (V_0(f_{0,\epsilon}) - V_0(f_0)). \end{aligned}$$

If $P_0 \in \mathcal{M}^J$, we have $\sup_{f\in\mathcal{F}} \mid V_0(f,\epsilon) - V_0(f) \mid = O(\epsilon^J)$. Thus,

$$\begin{aligned} V_0(f_{n,B_n},\epsilon) - V_0(f_{n,B_n}) &= O(\epsilon^J) \\ V_0(f_{0,\epsilon},\epsilon) - V_0(f_{0,\epsilon}) &= O(\epsilon^J). \end{aligned}$$

In addition, we have

$$\begin{aligned} 0 &\leq V_0(f_{0,\epsilon}) - V_0(f_0) \\ &= (V_0(f_{0,\epsilon}) - V_0(f_{0,\epsilon},\epsilon)) + (V_0(f_0,\epsilon) - V_0(f_0)) \\ &\quad + (V_0(f_{0,\epsilon},\epsilon) - V_0(f_0,\epsilon) \\ &\leq (V_0(f_{0,\epsilon}) - V_0(f_{0,\epsilon},\epsilon)) + (V_0(f_0,\epsilon) - V_0(f_0)), \end{aligned}$$

by using that $V_0(f_{0,\epsilon}) - V_0(f_0) \leq 0$ since $f_0$ is the minimizer of $V_0(f)$. Thus, the proves that also $V_0(f_{0,\epsilon}) - V_0(f_0) = O(\epsilon^J)$. This completes the proof of the theorem. $\Box$

## 6.4 Oracle inequality

Theorems 4 and 5 establish the desired oracle inequality.

**Theorem 6** *Let $P_0 \in \mathcal{M}(J)$. Assume $M_1 = \sup_{P\in\mathcal{M}(0)} \sup_{f\in\mathcal{F}} \sup_o \mid L_{\bar{Q},G}(f)(o) \mid < \infty$. Assume that $G_0$ is known and $G_n = G_0$. Let $m_n$ be either the CV-TMLE, CV-AIPTW, CV-IPTW selector. Then,*

$$\begin{aligned} Ed_0(\hat{f}_{m_n}(P^0_{n,B_n}), f_0) &\leq E \min_m E_{B_n} d_0(\hat{f}_m(P^0_{n,B_n}), f_0) \\ &\quad + C(M_1) \frac{\sqrt{\log M_n + \log n}}{(n\epsilon p)^{0.5}} + O(\epsilon^J). \end{aligned}$$

*In the case that $m_n$ equals $m_{n,IPTW}$ or $m_{n,AIPTW}$ one can remove the $\log n$ in the remainder term.*

*If $\epsilon_n = n^{-J/(2J+1)}$, then we obtain*

$$\begin{aligned} Ed_0(\hat{f}_{m_n}(P^0_{n,B_n}), f_0) &\leq E \min_m E_{B_n} d_0(\hat{f}_m(P^0_{n,B_n}), f_0) \\ &\quad + C(M_1) \frac{\sqrt{\log M_n + \log n}}{n^{J/(2J+1)}}. \end{aligned}$$

## 6.5 Asymptotic equivalence of cross-validation selector and the oracle selector.

In the typical scenario that the oracle selected estimator converges to the truth $f_0$ at a slower rate than $n^{-J/(2J+1)}$, it is reasonable to assume

$$(\log M_n + \log n)^{0.5} \frac{n^{-J/(2J+1)}}{E \min_m E_{B_n} d_0(\hat{f}_m(P^0_{n,B_n}), f_0)} \rightarrow 0.$$

Dividing both sides of the above finite sample oracle inequality by the latter denominator yields:

$$\frac{Ed_0(\hat{f}_{m_n}(P^0_{n,B_n}), f_0)}{E \min_m E_{B_n} d_0(\hat{f}_m(P^0_{n,B_n}), f_0)} \rightarrow 1.$$

In other words, in this typical scenario the discrete super-learner is asymptotically equivalent with the oracle selected estimator.

# 7 An estimator achieving minimax rate driven by size of parameter space of optimal rule

van der Laan, Dudoit (2003) and van der Laan et al. (2006) proposed an $\delta$-net estimator that achieves the minimax rate defined by the entropy of the parameter space, and developed a finite sample inequality of the risk based dissimilarity for this estimator expressed by subsequent application of the oracle inequality for the cross-validation selector. Here we apply this estimator and method of proof to our case for fixed $\epsilon$ to establish analogue results.

**Theorem 7** *Let $\mathcal{F}$ be a parameter space for $f_0$. Let $N(\delta, \mathcal{F}, \| \cdot \|)$ be the $\delta$-covering number of $\mathcal{F}$ w.r.t. a norm so that $d_0(f, f_0) = O(\| f - f_0 \|)$. Assume $\sqrt{\log N(\delta, \mathcal{F}, \| \cdot \|)} = O(\delta^{-(1-\alpha)})$. Let $\mathcal{F}_\delta = \{f_1^\delta, \ldots, f_{N(\delta)}^\delta\}$ be such a $\delta$-covering so that $N(\delta) = O(N(\delta, \mathcal{F}, \| \cdot \|))$. Define the estimator*

$$\hat{f}_{\delta,\epsilon}(P_n) = \arg\min_{f \in \mathcal{F}_\delta} V_n^*(f, \epsilon).$$

*Let*

$$\delta_n = \arg\min_{\delta} V_n^*(\hat{f}_\delta, \epsilon)$$

*be the CV-TMLE selector for the library of candidate estimators $\{\hat{f}_{\delta,\epsilon} : \delta\}$. We have*

$$E_0 d_0(\hat{f}_{\delta_n,\epsilon}(P^0_{n,B_n}), f_0) = O\left((n\epsilon)^{-1/(4-2\alpha)} + \epsilon^J\right).$$

*If we select*

$$\epsilon_n = n^{-m(\alpha,J)}, \text{ where } m(\alpha, J) = \tfrac{1/(2-\alpha)}{2J+1/(2-\alpha)},$$

*then*

$$E_0 d_0(\hat{f}_{\delta_n,\epsilon}(P^0_{n,B_n}), f_0) = O\left(n^{-m_1(\alpha,J)}\right), \text{ where } m_1(\alpha, J) = \tfrac{J}{4J-2J\alpha+1}. \quad (15)$$

If $\mathcal{F}$ is the class of cadlag function on $[0, \tau] \subset \mathbb{R}^d$ with variation norm bounded by some universal constant $M$, then we have $\alpha = \alpha(d) = 1/(d+1)$. Inclusion of this estimator in the library of the super-learner guarantees this minimal rate (15) of convergence for the super-learner. The presented rate shows that even for this highly nonparametric class of possible rules we can achieve a rate close to $n^{-1/4}$.

This $\delta$-net estimator was proposed in van der Laan, Dudoit (2003) and van der Laan et al (2006). As shown in these papers, we can also define such $\delta$-nets for a large collection of subspaces of the most nonparametric set $\mathcal{F}$, and also select the choice of subspace with the CV-TMLE selector (i.e., $\delta_n$ above is replaced by joint index for both the resolution and the choice of subspace). In that case, we would achieve an adaptive rate that adapts to the size of the smallest subspace containing the true $f_0$. It is straightforward to generalize Theorem 7 to this more general and more adaptive $\delta$-net estimator.

Alternatively, we note that the estimator presented in this theorem is indexed by a choice of subspace $\mathcal{F}$, and we can add a whole collection of such $\mathcal{F}$-specific estimators to the library of the super-learner for which we have an oracle inequality demonstrating that we will do as well as the best choice of estimator in the library, under our stated assumptions.

**Proof:** We can represent the estimator $\hat{f}_{\delta,\epsilon}(P_n)$ as a cross-validation selector as follows. Let $\hat{f}_j^\delta(P_n) = f_j^\delta$, $j = 1, \ldots, N(\delta)$. The cross-validation selector is defined by

$$j_n = \arg\min_j E_{B_n} V_n^*(\hat{f}_j^\delta(P_{n,B_n}^0), \epsilon_n) = \arg\min_j V_n^*(f_j^\delta).$$

Now note that $\hat{f}_\delta(P_n) = \hat{f}_{j_n}^\delta(P_n)$ is indeed the cross-validated selected estimator for this library of constant estimators.

Applying the finite sample oracle inequality yields

$$E_0 E_{B_n} d_0(\hat{f}_\delta(P_{n,B_n}^0), f_0) \leq E_0 \min_j d_0(f_j^\delta, f_0) + C\left\{\frac{\sqrt{\log N(\delta)}}{n\epsilon} + \epsilon^J\right\}.$$

Applying the oracle inequality to the CV-TMLE selector $\delta_n$ over a finite grid with $K_n$ values yields:

$$E_0 d_0(\hat{f}_{\delta_n}, f_0) \leq E_0 \min_\delta E_{B_n} d_0(\hat{f}_\delta(P_{n,B_n}^0), f_0) + C\left\{\frac{\sqrt{\log K_n}}{n\epsilon} + \epsilon^J\right\},$$

where $K_n$ can be chosen to be $O(n)$, so that $\log K_n$ behaves as $\log n$. Combining the two inequalities yields:

$$\begin{aligned}
E_0 d_0(\hat{f}_{\delta_n}, f_0) &\leq E_0 \min_\delta E_{B_n} d_0(\hat{f}_\delta(P_{n,B_n}^0), f_0) + C\left\{\frac{\sqrt{\log K_n}}{n\epsilon} + \epsilon^J\right\} \\
&\leq E_0 \min_\delta \left\{E_0 \min_j d_0(f_j^\delta, f_0) + C\left\{\frac{\sqrt{\log N(\delta)}}{n\epsilon} + \epsilon^J\right\}\right\} \\
&\quad + C\left\{\frac{\sqrt{\log K_n}}{n\epsilon} + \epsilon^J\right\} \\
&\leq C E_0 \min_\delta \left\{\delta + \left\{\frac{\sqrt{\log N(\delta)}}{n\epsilon} + \epsilon^J\right\}\right\} \\
&\quad + C\left\{\frac{\sqrt{\log K_n}}{n\epsilon} + \epsilon^J\right\},
\end{aligned}$$

where we used that $d_0(f_j^\delta, f_0) = O(\delta)$. The optimal $\delta = (n\epsilon)^{-1/(4-2\alpha)}$. So the right-hand side behaves as $(n\epsilon)^{-1/(4-2\alpha)} + \epsilon^J$. The optimal rate $\epsilon$ is thus given by $\epsilon_n$ as stated in theorem. For this rate we obtain the rate $n^{-m_1(\alpha,J)}$. This completes the proof. $\Box$

# 8 Discussion

In this article we developed a highly adaptive and efficient estimator $V_n^*(\hat{f}, \epsilon_n)$ of counterfactual mean outcome $V_0(\hat{f})$ for a given candidate estimator $\hat{f}$ of the optimal rule $f_0$. Our estimator is a CV-TMLE of an $\epsilon$-approximation $V_0(\hat{f}, \epsilon)$ and $\epsilon$ is optimally selected with our data adaptive selector. This estimator achieves the rate $n^{-J_0/(2J_0+1)}$, where $J_0$ is the underlying unknown smoothness of $a \rightarrow \bar{Q}_0(a, X)$. In addition, we have shown that our estimator converges at this rate to a normal limit distribution, allowing the construction of a confidence interval. These results also provide us with state of the art inference for the counterfactual mean $EY_{g^*}$ for any static, dynamic or stochastic intervention $g^*$. This estimator of $V_0(\hat{f})$ provides the basis for estimation of the optimal treatment allocation rule $f_0$. Specifically, we build a library of candidate estimators of $f_0$, and then select the estimator that minimizes the estimated risk $V_n^*(\hat{f}, \epsilon_n)$ over all these candidate estimators. We establish an oracle inequality for this super-learner of $f_0$ which demonstrates that under weak conditions this super-learner of $f_0$ will be asymptotically equivalent with the oracle selector w.r.t. the dissimilarity $V_0(\hat{f}) - V_0(f_0)$. That is, it will perform optimally in estimating the optimal rule w.r.t the criterion that matters, namely, the performance of the individualized treatment rule when applied to the target population. We also propose a particular $\delta$-net estimator that achieves a rate implied by the entropy of the chosen parameter space for the optimal rule. Including these $\delta$-net estimators for a range of subspaces of a nonparametric function space yields a highly adaptive super-learner of $f_0$ that is guaranteed to converge at a specified minimal rate close to $n^{-1/4}$, but potentially might converge much faster. Finally, we note that our approach for constructing an estimator and inference for a non-pathwise differentiable target parameter $f_0$ is completely general and can be applied to a large class of non-pathwise differentiable target parameters, including our selector of $\epsilon$.

## Acknowledgement

This grant is funded by NIH-grant 5R01AI074345-07.

# Appendix

## A Oracle inequality for quadratic loss-based dissimilarity

In this section we establish a finite sample oracle inequality and asymptotic equivalence result for the special case that $P_0 \in \mathcal{M}_I(J)$ for some $J \geq 2$. We first make the following observation.

**Lemma 4** *Assume $P_0 \in \mathcal{M}_I(J)$, $J \geq 2$. Let $M_1(\epsilon) = \sup_{o,f,\bar{Q}} \mid L_{\bar{Q},G_0,\epsilon}(f)(o) - L_{\bar{Q},G_0,\epsilon}(f_{0,\epsilon}) \mid$ be the uniform norm of $L_{\bar{Q},G_0,\epsilon}(f)(O)$ over all possible $f$, $O$ over a support of $P_0$, and $\bar{Q}$ over the parameter space of $\bar{Q}_0$. Let $M_2(\epsilon)$ be such that uniformly in $f$ and $\bar{Q}$*

$$P_0(L_{\bar{Q},G_0,\epsilon}(f) - L_{\bar{Q},G_0,\epsilon}(f_{0,\epsilon}))^2 \leq M_2(\epsilon) P_0(L_{\bar{Q},G_0,\epsilon}(f) - L_{\bar{Q},G_0,\epsilon}(f_{0,\epsilon})).$$

*We have*

$$\begin{aligned} M_1(\epsilon) &= O(\epsilon^{-1}) \\ M_2(\epsilon) &= O(\epsilon^{-3}). \end{aligned}$$

**Proof:** It is straightforward to show $M_1(\epsilon) = O(\epsilon^{-1})$. Let's consider the IPCW-loss function: $L_{G_0,\epsilon}(f) = \frac{k_{f(X),\epsilon}(A)Y}{g_0(A|X)}$. Let $\sigma^2(a,w) = E_0(Y^2 \mid A = a, W = w)$ and let $\xi(y,X) = \xi(y, y + (f - f_0)(X)/\epsilon)$ be so that $k(y + (f - f_0)(X)/\epsilon) - k(y) = k^{(1)}(\xi)(f - f_0)(X)/\epsilon$. We have

$$\begin{aligned}
&P_0\{L_{G_0,\epsilon}(f) - L_{G_0,\epsilon}(f_{0,\epsilon})\}^2 = E_{P_0} \int_a \frac{\sigma_0^2(a,X)}{g_0(a|X)} \left\{k_{f(X),\epsilon}(a) - k_{f_{0,\epsilon}(X),\epsilon}(a)\right\}^2 da \\
&= \tfrac{1}{\epsilon} E_{P_0} \int_a \frac{\sigma_0^2(f(X)+\epsilon y, X)}{g_0(f(X)+\epsilon y|X)} \left\{k(y + (f - f_0)(X)/\epsilon) - k(y)\right\}^2 dy \\
&= \tfrac{1}{\epsilon^3} E_{P_0} \int_a \frac{\sigma_0^2(f(X)+\epsilon y, X)}{g_0(f(X)+\epsilon y|X)} \left\{k^{(1)}(\xi(y,x))\right\}^2 (f - f_0)^2(X) dy \\
&= O(\epsilon^{-3} \parallel f - f_0 \parallel_{P_0}^2).
\end{aligned}$$

We also have

$$\begin{aligned}
P_0\{L_{G_0,\epsilon}(f) - L_{G_0,\epsilon}(f_{0,\epsilon})\} &= E_{P_0} \int_a \{\bar{Q}_{0,\epsilon}(f(X),X) - \bar{Q}_{0,\epsilon}(f_{0,\epsilon}(X),X)\} \\
&= 0.5 E_{P_0} \int_a \bar{Q}_{0,\epsilon}^{(2)}(\xi(f, f_{0,\epsilon}, X))(f - f_{0,\epsilon})^2(X).
\end{aligned}$$

Thus, if $\inf_{a \in [0,1]} \bar{Q}_0^{(2)}(a,X) > \delta > 0$ for some $\delta > 0$, $P_{0,X}$-a.e, then it follows that indeed $M_2(\epsilon) = O(\epsilon^{-3})$. $\square$

This allows us to apply the general oracle inequalities in van der Laan, Dudoit (2003) for $m_{n,AIPTW} = \arg\min_m E_{B_n} P^1_{n,B_n} L_{Q_{n,B_n},G_{n,B_n},\epsilon}(\hat{f}_m(P^0_{n,B_n}))$, and thus also for $m_{n,IPTW}$. However, since $G_0$ is known we can rely on the simpler results in van der Laan and Dudoit (2003) and van der Vaart et a. (2006) for fixed loss functions $L(f)$ (not indexed by a nuisance parameter). For $m_{n,IPTW}$ one can immediately apply these results since it relies on a fixed loss function $L_{G_0,\epsilon}$ (i.e., no nuisance parameter). Regarding, $m_{n,AIPTW}$, we note that the proofs in these article are trivially generalized to the case that the loss function is indexed by a $\bar{Q}$ that does not affect its risk, when it is estimated based on training sample : i.e. $P_0 L_{\bar{Q},G_{0,\epsilon}}(f) = V_0(f,\epsilon)$ for any $\bar{Q}$. So we can conclude that oracle inequalities in these articles immediately apply to the IPTW and augmented-IPTW cross-validation selectors $m_{n,IPTW}, m_{n,AIPTW}$.

On the other hand, the CV-TMLE cross-validation selector uses $\bar{Q}_{n,B_n,\epsilon_n}$ as an estimator of $\bar{Q}_0$, so that this estimator also depends on the validation sample through $\epsilon_n$. This is a minor hurdle in the general proof of the oracle inequality since empirical means $(P_n - P_0)f_{\epsilon_n}$ with random functions $f_{\epsilon_n}$ only random through a low dimensional random vector $\epsilon_n$ are easily handled by empirical process theory. Therefore, without proof, we state that the inequality is generalized easily by adding a multiplication factor $\log n$ to the remainder term. This results in the following theorem.

**Theorem 8** *For each $\delta > 0$, for each $\epsilon > 0$, we have*

$$\begin{aligned} Ed_{0,\epsilon}(\hat{f}_{m_n}(P^0_{n,B_n}), f_{0,\epsilon}) &\leq (1+2\delta) E \min_m E_{B_n} d_{0,\epsilon}(\hat{f}_m(P^0_{n,B_n}), f_{0,\epsilon}) \\ &\quad + C(\delta)\frac{\log M_n + \log n}{\epsilon^3 np}, \end{aligned}$$

*where $C(\delta) = O(\delta^{-1})$. In the special case that $m_n$ equals $m_{n,IPTW}$ or $m_{n,AIPTW}$, the remainder term can be chosen to be $C(M_1(\epsilon), M_2(\epsilon), \delta)\log M_n/(np)$, where $C(M_1, M_2, \delta) = 2(1+\delta)^2(2M_1/3 + M_2/\delta)$.*

In order to move towards a result in terms of $d_0()$, we establish the following lemma.

**Lemma 5** *Assume $P_0 \in \mathcal{M}_I(J)$, $J \geq 3$. We have*

$$d_{0,\epsilon}(f, f_{0,\epsilon}) - d_0(f, f_0) = O\left(\| f_{0,\epsilon} - f_0 \|_\infty + \frac{\| f_{0,\epsilon} - f_0 \|}{\| f - f_0 \|}\right) d_0(f, f_0)).$$

**Proof:** We have

$$\begin{aligned} d_{0,\epsilon}(f, f_{0,\epsilon}) - d_0(f, f_0) &= 0.5 E_{P_0}\left\{\bar{Q}^{(2)}_{0,\epsilon}(\xi)(f - f_{0,\epsilon})^2(X) - \bar{Q}^{(2)}_{0,\epsilon}(\xi_1)(f - f_0)^2(X)\right\} \\ &= 0.5 E_{P_0}(\bar{Q}^{(2)}_{0,\epsilon}(\xi) - \bar{Q}^{(2)}_{0,\epsilon}(\xi_1))(f - f_0)^2 \\ &\quad + 0.5 E_{P_0}\bar{Q}^{(2)}_{0,\epsilon}(\xi)\left\{(f - f_{0,\epsilon})^2(X) - (f - f_0)^2(X)\right\}, \end{aligned}$$

where $\xi = \xi(\epsilon, X)$ is a value between $f(X)$ and $f_{0,\epsilon}(X))$, while $\xi_1 = \xi_1(\epsilon, X)$ is a value between $f(X)$ and $f_0(X)$. Thus, $\xi - \xi_1 = O(\mid f_0 - f_{0,\epsilon} \mid)$, so that, if $\bar{Q}_{0,\epsilon}^{(3}$ exists and its continuous, then the first term is $O\left(\| f_{0,\epsilon} - f_0 \|_\infty\right) d_0(f, f_0)$. Here we use that for $P_0 \in \mathcal{M}_I(J)$ the dissimilarities $d_0$ and $d_{0,\epsilon}$ are equivalent with $\| \cdot \|_{P_0}^2$. For the second term we note that

$$
\begin{aligned}
(f - f_{0,\epsilon})^2 - (f - f_0)^2 &= (f_{0,\epsilon} - f_0)(2f - f_0 - f_{0,\epsilon}) \\
&= (f_{0,\epsilon} - f_0)(2(f - f_0) - (f_{0,\epsilon} - f_0)) \\
&= -(f_{0,\epsilon} - f_0)^2 + 2(f_{0,\epsilon} - f_0)(f - f_0).
\end{aligned}
$$

Thus, the second term is $O(\| f_{0,\epsilon} - f_0 \| / \| f - f_0 \|) d_0(f, f_0))$. To conclude,

$$
d_{0,\epsilon}(f, f_{0,\epsilon}) - d_0(f, f_0) = O\left(\| f_{0,\epsilon} - f_0 \|_\infty + \frac{\| f_{0,\epsilon} - f_0 \|}{\| f - f_0 \|}\right) d_0(f, f_0)).
$$

This completes the proof. □

We note that if $P_0 \in \mathcal{M}_I(J)$, $J \geq 2$, it is not hard to establish that $f_{0,\epsilon} - f_0 = O(\epsilon^J)$ w.r.t. supremum norm and thus also w.r.t. $L^2(P_0)$-norm. In combination with the previous lemma this results in the following theorem providing conditions under which the cross-validation selected estimator is asymptotically equivalent with the oracle selector.

**Theorem 9** *Assume $P_0 \in \mathcal{M}_I(J)$, $J \geq 3$. We have $\| f_{0,\epsilon} - f_0 \|_\infty = O(\epsilon^J)$. Let $\epsilon_n$ be such that*

$$
\begin{aligned}
\frac{\epsilon^J}{E \min_m E_{B_n} d_{0,\epsilon}(\hat{f}_m(P_{n,B_n}^0), f_{0,\epsilon})} &= o(1) \\
\frac{\epsilon^J}{(\log M_n + \log n)\epsilon^{-3} n^{-1}} &= o(1).
\end{aligned}
$$

*Then,*

$$
\frac{E d_0(\hat{f}_{m_n}(P_{n,B_n}^0), f_0)}{E \min_m E_{B_n} d_0(\hat{f}_m(P_{n,B_n}^0), f_0)} \to 1 \ \textit{as}\ n \to \infty.
$$