Chapter 1
Targeted Learning of Optimal Individualized Treatment Rules Under Cost Constraints

Boriska Toth and Mark van der Laan

1.1 Introduction

We consider a general resource-allocation problem, namely, to maximize a mean outcome given a cost constraint, through the choice of a treatment rule that is a function of an arbitrary fixed subset of an individual's covariates. In pharmaceutical applications, we typically think of maximizing a clinical outcome given a monetary cost constraint, through the allocation of medication to patients, although our model is much more general. We focus on the setting where unmeasured confounding is a possibility, but a valid instrumental variable is available. Thus, our setup allows for consistent estimation of the optimal treatment rule and causal effects in a range of non-randomized studies, including post-market and other observational studies, as well as studies involving imperfect randomization due to non-adherence. The goal is both to: (1) find an optimal intervention $d(V)$ for maximizing the mean counterfactual outcome, where V is an arbitrary fixed subset of baseline covariates W, and (2) estimate the mean counterfactual outcome under this rule $d(V)$. We make no restrictions on the type of data; however, the case of a continuous or categorical instrument or treatment variable is discussed in Toth (2016). To our knowledge, this work is the first to estimate the effect of an optimal individualized treatment regime, under a non-unit cost constraint, in the instrumental variables setting.

Utilizing instrumental variables. A classic solution for obtaining a consistent estimate of a causal effect under unmeasured confounding is to use an instrumental variable, assuming one exists. Informally, an instrumental variable, or instrument, is a variable Z that affects the outcome Y only through its effect on the treatment A, and the residual (error) term of the instrument is uncorrelated with the residual term of the

B. Toth (✉) · M. van der Laan
UC-Berkeley, Berkeley, USA
e-mail: bori@stat.berkeley.edu

M. van der Laan
e-mail: laan@berkeley.edu

K. E. Peace et al. (eds.), *Biopharmaceutical Applied Statistics Symposium*,
ICSA Book Series in Statistics, https://doi.org/10.1007/978-981-10-7820-0_1

outcome (Imbens and Angrist 1994; Angrist et al. 1996; Angrist and Krueger 1991). Thus, the instrument produces exogenous variation in the treatment. Instrumental variables have been used widely in biostatistics and pharmaceutics. (See Brookhart et al. 2010 for a large collection of references.) In these settings, the instrumental variable is usually some attribute that is related to the health care a patient receives, but is not at the level of individual patients. For example, Brookhart and Schneeweiss (2007) exploit variation in physician preference for prescribing NSAID medications to infer the effect of these medications on gastrointestinal bleeding.

In this work, we solve two versions of the optimal individualized treatment problem: (1) when the intervention is on the treatment variable A (Sect. 1.7), and (2) when the intervention is actually on the instrument Z (Sect. 1.6). For example, consider a study in which HIV-positive patients were encouraged to undergo antiretroviral therapy (ART) with a randomized (or quasi-randomized) encouragement design, but a number of factors caused non-adherence among some patients (Chesney 2006). The methods in this chapter allow one to infer what would be the optimal assignment of patients to ART treatment, based on patient characteristics, to achieve a desirable outcome (i.e., suppressed viral load, 5-year survival), given a limited budget. One parameter of interest is the mean outcome under optimal assignment of individuals to actually receive ART. This is the problem of finding an optimal treatment regime. However, in this setting of non-adherence, it might not be possible to intervene directly on the treatment variable. Thus, another parameter of interest is the mean outcome under the optimal intervention on the instrumental variable. We call this the problem of finding an optimal *intent-to-treat* regime, so named because the instrument is often a randomized assignment to treatment or encouragement mechanism. Under our randomization assumption on instrument Z, the optimal intent-to-treat problem is the same as an optimal treatment problem without unmeasured confounding, as Z can be seen as a treatment variable that is unconfounded with Y.

Causal effects given arbitrary subgroups of the population.
A key feature of our work is that the optimal intervention $d(V)$ is a function of a fixed arbitrary subset V of all baseline covariates W. There is currently great interest and computational feasibility in designing individualized treatment regimes based on a patient's characteristics and biomarkers. The paradigm of precision medicine calls for incorporating high-dimensional spaces of genetic, environmental, and lifestyle variables into treatment decisions (Editors: National Research Council Committee 2011). Incorporating many covariates for estimating relevant components of the data-generating distribution can be helpful in: (1) improving the precision of the statistical model and (2) ensuring that the instrument induces exogenous variation given the covariates. However, a physician typically has a smaller set of patient variables that are available and that he/she considers reliable predictors. Thus, being able to calculate an optimal treatment (or intent-to-treat) regime as a function of an arbitrary subset of baseline covariates is of great use.

The targeted minimum loss-based framework.
Our estimators use targeted minimum loss-based estimation (TMLE), which is a methodology for semiparametric estimation that has very favorable theoretical

properties and can be superior to other estimators in practice (van der Laan and Rubin 2006; van der Laan and Rose 2011). TMLE guarantees asymptotic efficiency when certain components of the data-generating distribution are consistently estimated. Thus, under certain conditions, the TMLE estimator is optimal in having the asymptotically lowest variance for a consistent estimator in a general semiparametric model, thereby achieving the semiparametric Cramer–Rao lower bound (Newey 1990). The TMLE method also has a robustness guarantee: It produces consistent estimates even when the functional form is not known for all relevant components of the parameter of interest (see Sects. 1.6.3.4 and 1.7.3). Another beneficial property is asymptotic linearity. This ensures that TMLE-based estimates are close to normally distributed for moderate sample sizes, which makes for accurate coverage of confidence intervals. Finally, TMLE has the advantage over other semiparametric efficient estimators that it is a substitution estimator, meaning that the final estimate is made by evaluating the parameter of interest on the estimates of its relevant components. This property has been linked to good performance in sparse data in Gruber and van der Laan (2010).

The TMLE methodology uses the following procedure for constructing an estimator:

1. Let P_0 denote the true data-generating distribution. One first notes that the parameter of interest $\Psi(P_0)$ depends on P_0 only through certain relevant components Q_0 of the full distribution P_0; in other words, $\Psi(P_0) = \Psi(Q_0)$.[1] TMLE *targets* these relevant components by only estimating these Q_0 and certain nuisance parameters g_0[2] that are needed for updating the relevant components. An initial estimate (Q_n^0, g_n) is formed of the relevant components and nuisance parameters. This is typically done using the Super Learner approach described in van der Laan et al. (2007), in which the best combination of learning algorithms is chosen from a library using cross-validation.
2. Then, the relevant components Q_n^0 are fluctuated, possibly in an iterative process, in an optimal direction for removing bias efficiently. To do so, one defines a fluctuation function $\varepsilon \to Q(\varepsilon|g_n)$ and a loss function $L(\dots)$, where we fluctuate Q_n^0 to $Q_n^0(\varepsilon|g_n)$ by solving for fluctuation $\varepsilon = \text{argmin}_\varepsilon \frac{1}{n}\sum_{i=1}^n L(Q_n^0(\varepsilon|g_n), g_n)(O_i)$. For example, the loss function might be the mean squared error or the negative log likelihood function.
3. Finally, one evaluates the statistical target parameter on the updated relevant components Q_n^* and arrives at estimate $\psi_n^* = \Psi(Q_n^*)$.

The key requirement is to choose the fluctuation and loss functions so that, upon convergence of the components to their final estimate Q_n^* and g_n^*, the efficient influence curve equation is solved:

$$P_n\, D^*(Q_n^*, g_n^*) = 0$$

[1] We are abusing notation here for the sake of convenience by using $\Psi(\cdot)$ to denote the mapping both from the full distribution to $\mathbb{R}^d$ and from the relevant components to $\mathbb{R}^d$.

[2] The nuisance parameters are those components g_0 of the efficient influence curve $D^*(Q_0, g_0)$ that $\Psi(Q_0)$ does not depend on.

Above, P_n denotes the empirical distribution $(O_1, \ldots, O_n)$, and we use the shorthand notation $P_n f = \frac{1}{n}\sum_{i=1}^{n} f(O_i) \cdot D^*$ denotes the efficient influence curve.

1.2 Prior Work

Luedtke and van der Laan (2016a) is a recent work that gives a TMLE estimator for the mean outcome under optimal treatment given a cost constraint. That problem is very similar to the one we solve in Sect. 1.6, with the main difference being that we allow a more general non-unit cost constraint which results in a different closed-form solution to the optimal rule. Luedtke and van der Laan (2016b) tackles the issue of possible non-unique solutions and resulting violations of pathwise differentiability. The conditions we require in assumptions (A2)–(A4) are adopted from these works.

A large body of work focuses on the case of optimal treatment regimes in the unconstrained case, such as Robins (2004). More recently, various approaches tackle the constrained ODT problem: Zhang et al. (2012) describe a solution that assumes the optimal treatment regime is indexed by a finite-dimensional parameter, while Chakraborty et al. (2013) describe a bootstrapping method for learning ODT regimes with confidence intervals that shrink at a slower than root-n rate. Chakraborty and Moodie (2013) give a review of recent work on the constrained case.

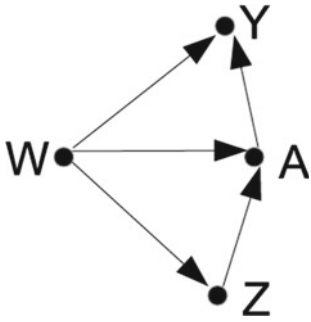
1.3 Model and Problem

We consider the problem of estimation and inference under an optimal intervention, in the context of an instrumental variable model. We take an iid sample of n data points $(W, Z, A, Y) \sim \mathscr{M}$, where $\mathscr{M}$ is a semiparametric model. Z is assumed to be a valid instrument for identifying the effect of treatment A on outcome Y, when one has to account for unmeasured confounding. In applications, instrument Z is often a randomized encouragement mechanism or randomized assignment to treatment which may or may not be followed. In other cases, Z is not perfectly randomized but nevertheless promotes or discourages individuals in receiving treatment. $V \subseteq W$ is an arbitrary fixed subset of the baseline covariates, and $F_V(W)$ gives the mapping $W \to V \cdot d(V)$ refers to a decision rule as a function of V, where $Z = d(V)$ is used to denote the optimal intervention on the instrument Z, in other words, the optimal assignment to treatment or the optimal intent-to-treat. $A = d(V)$ refers to the optimal treatment rule. We are interested in estimating the mean counterfactual outcome under an optimal rule $Z = d(V)$ or $A = d(V)$. Figure 1.1 shows a diagram.

There are no restrictions on the type of data. However, the case of categorical or continuous Z or A are both dealt with separately in Toth (2016).

Further, we let $c_A(A, W)$ be a cost function that gives the cost associated with assigning an individual with covariates W to a particular A value. We let $c_T(Z, W)$ be a cost function that gives the total cost associated with assigning an individual

Fig. 1.1 Causal diagram



with covariates W to a particular Z value. We can think of $c_T(Z, W)$ as the sum of $c_Z(Z, W)$, a cost incurred directly from setting Z, and $E_{A|W,Z}c_A(A, W)$, an average cost incurred from the actual treatment A.[3] We need to find optimal rule $Z = d(V)$ under cost constraint $E\ c_T(Z, W) \leq K$, for a fixed cost K, and optimal rule $Z = d(V)$ under constraint $E\ c_A(A, W) \leq K$.

Notation. Let $P_W \equiv \Pr(W)$ and $\rho(Z, W) \equiv \Pr(Z = 1|W)$. Also let $\Pi(Z, W) \equiv E(A \mid Z, W)$ be the conditional mean of A given Z, W, and $\mu(Z, W) \equiv E(Y \mid Z, W)$.

We also define $\mu_b(V) \triangleq E_{W|V}\big[\mu(Z = 1, W) - \mu(Z = 0, W)\big]$, which gives the mean difference in outcome between setting $Z = 1$ and $Z = 0$ given V. Similarly, $c_{b,Z}(V) \triangleq E_{W|V}\big[c_T(Z = 1, W) - c_T(Z = 0, W)\big]$, and $c_{b,A}(V) \triangleq E_{W|V}\big[c_A(A = 1, W) - c_A(A = 0, W)\big]$. We also use notation $m(V) \triangleq E_{W|V}m(W)$, where m is the causal effect function defined in the causal assumptions.

We further assume wlog that intent-to-treat $Z = 0$ has lower cost for all V: $E_{W|V}c_T(0, W) \leq E_{W|V}c_T(1, W)$.[4] Let $\underline{K_Z} \triangleq E_W c_T(0, W)$ be the total cost of not assigning any individuals to intent-to-treat, and $\overline{K_Z} \triangleq E_W c_T(1, W)$ be the total cost of assigning everyone, and we assume a non-trivial constraint $\underline{K_Z} < K < \overline{K_Z}$. Define $\underline{K_A} \triangleq E_W c_A(0, W)$, and $\overline{K_A}$ similarly.

Causal model.
Using the structural equation framework of (Pearl 2000), we assume that each variable is a function of other variables that affect it and a random term (also called error term). Let U denote the error terms. Thus, we have

$$W = f_W(U_W), Z = f_Z(W, U_Z), A = f_A(W, Z, U_A), Y = f_Y(W, Z, A, U_Y)$$

[3]It is not hard to extend this model to incorporate uncertainty in $E(A|W, Z)$ for calculating $c_T(Z, W)$, and thus estimating $c_T(Z, W)$ from the data, given fixed functions c_Z, c_A. There is a correction term that gets added to the efficient influence curve.

[4]We are only making this assumption for the sake of easing notation. We can forgo this assumption by introducing notation; i.e., $Z = l(V)$ is the lower cost intent-to-treat value for a stratum defined by covariates V.

where $U = (U_W, U_Z, U_A, U_Y) \sim P_{U,0}$ is an exogenous random variable, and f_W, f_Z, f_A, f_Y may be unspecified or partially specified (for instance, we might know that the instrument is randomized). U_Y is possibly confounded with U_A.

We use notation that a subscript of 0 denotes the true distribution, in expressions such as E_0, P_0.

Assumption (A1) ***Assumptions ensuring that Z is a valid instrument***:

1. ***Exclusion restriction***. Z only affects outcome Y through its effect on treatment A. Thus, $f_Y(W, Z, A, U_Y) = f_Y(W, A, U_Y)$.
2. ***Exogeneity of the instrument***. $E(U_Y|W, Z) = 0$ for any W, Z.
3. Z ***induces variation in*** A. $\text{Var}_0[E_0(A|Z, W)|W] > 0$ for all W.

Structural equation for outcome Y:

4. $Y = Am(W) + \theta(W) + U_Y$ for continuous Y, and
   $\Pr(Y = 1|W, A, \tilde{U}_Y) = Am(W) + \theta(W) + \tilde{U}_Y$ for binary Y,
   where $U_Y = (\tilde{U}_Y, U'_Y)$ for an exogenous r.v. U'_Y,[5] and m, θ are unspecified functions.

Assumptions 2 and 4 yield that, whether Y is binary or continuous,

$$E(Y|W, Z) = m_0(W)\Pi_0(W, Z) + \theta_0(W)$$

We use $Y(A = a)$ to denote the counterfactual from setting treatment to $A = a$. These assumptions guarantee that $E(Y(A = a))$ equals $E_W m(W)a + \theta(W)$ for identifiable functions m, θ.

It should be noted that we do not require the instrument to be randomized with respect to treatment ($U_Z \perp\!\!\!\perp U_A|\ W$ is not necessary).

It is simple to see from the above instrumental variable assumptions that Z is randomized with respect to Y, so we have:

Corollary 1 (*Randomization of Z.*) $U_Z \perp U_Y|W$.

This implies $E(Y(Z)|W) = E(Y|W, Z)$.

Statistical model. The above-stated causal model implies the statistical model $\mathcal{M}$ consisting of all distributions P of $O = (W, Z, A, Y)$ satisfying $E_P(Y|W, Z) = m_P(W) \cdot \Pi_P(W, Z) + \theta_P(W)$. Here, m_P and θ_P are unspecified functions and $\Pi_P(W, Z) = E_P(A|W, Z)$ such that $Var_P(\Pi_P(Z, W)|W) > 0$ for all W. Note that the regression equation $E_P(Y|W, Z) = m_P(W) \cdot \Pi_P(W, Z) + \theta_P(W)$ is always satisfied for some choice of $m(W)$, $\theta(W)$ when Z is binary. The distribution for the instrument $\rho(W)$ may or may not be known, and we generally think of all other components P_W, Π, m, θ as unspecified.

[5] The U'_Y term is an exogenous r.v. whose purpose is for sampling binary Y with mean $\tilde{f}_Y(W, Z, A, \tilde{U}_Y)$.

1.3.1 Parameter of Interest, with Optimal Intent-to-Treat

Causal parameter of interest.

$$\Psi_Z(P_0) \triangleq \text{Max}_d \; E_{P_0} Y(Z = d(V)) \text{ s.t. } E_{P_0}[c_T(Z = d(V), W)] \leq K$$

Statistical target parameter.

$$\Psi_{Z,0} = E_{P_0}\mu_0(Z = d_0(V), W) \tag{1.1}$$

where d_0 is the optimal intent-to-treat rule:
$d_0 = \text{argmax}_d \; E_{P_0}\mu_0(Z = d(V), W)$ s.t. $E_{P_0}[c_T(Z = d(V), W)] \leq K$
We also use the notation $\Psi_Z(P_0) = \Psi_Z(P_{W,0}, \mu_0)$.

1.3.2 Parameter of Interest, with Optimal Treatment

Causal parameter of interest.

$$\Psi_A(P_0) \triangleq \text{Max}_d \; E_0 Y(A = d(V)) \text{ s.t. } E_0[c_A(A = d(V), W)] \leq K \tag{1.2}$$

Identifiability. $m(W)$ is identified as $\big[(\mu(Z = 1, W) - \mu(Z = 0, W))/(\Pi(Z = 1, W) - \Pi(Z = 0, W))\big]$. $\theta(W)$ is identified as $\big[\mu(Z, W) - \Pi(Z, W) \cdot m(W)\big]$.

Statistical target parameter.

Lemma 1 *The causal parameter given in Eq. (1.2) is identified by the statistical target parameter:*

$$\Psi_{A,0} = E_{P_{W,0}}\big[m_0(W)d_0(V) + \theta_0(W)\big] \tag{1.3}$$

Note that optimal decision rule d_0 is a function of m_0, $P_{W,0}$. For $\Psi_{A,0}$ we also use the notation $\Psi_A(P_{W,0}, m_0, \theta_0)$, or alternately $\Psi_A(P_{W,0}, \Pi_0, \mu_0)$, using the above identifiability results.

This lemma follows from our causal assumptions:

$\Psi_A(P_0) = EY(A = d_0(V)) = E_W E_{U_Y|W} EY(A = d_0(V)|W, U_Y)$
The right hand side becomes $E_W E_{U_W|Y}(m(W)d_0(V) + \theta(W) + U_Y)$ for a continuous Y, and $E_W E_{U_W|Y}(m(W)d_0(V) + \theta(W) + \tilde{U}_Y)$ for a binary Y.

1.4 Closed-Form Solution for Optimal Rule d_0 in the Case of Binary Treatment

The problem of finding the optimal deterministic treatment rule $d(V)$ is NP-hard (Karp 1972). However, when allowing possible non-deterministic treatments, there is a simple closed-form solution for the optimal treatment or the optimal intent-to-treat. The optimal rule is to treat all strata with the highest marginal gain per marginal cost, so that the total cost of the policy equals the cost constraint.

This section introduces key quantities and notation used in the rest of the chapter. We present the solution in detail for the case of intervening on the instrument, when $Z = d_0(V)$. Recall that wlog we think of $Z = 0$ as the 'baseline' intent-to-treat (ITT) value having lower cost. We define a scoring function $T(V) = \frac{\mu_b(V)}{c_b(V)}$ for ordering subgroups (given by V) based on the effect of setting $Z = 1$ per unit cost. In the optimal intent-to-treat policy, all groups with the highest $T(V)$ values deterministically have Z set to 1, up to cost K and assuming $\mu_b \geq 0$. We write $T_P(V)$ to make explicit the dependence on P_W, $\mu(Z, W)$ from distribution P.

Define a function S_P: $[-\infty, +\infty] \to \mathbb{R}$ as

$$S_P(x) = E_V[I(T_P(V) \geq x)(c_b(V)]$$

In other words, $S_P(x)$ gives the expected (additional above baseline) cost of setting $Z = 1$ for all subgroups having $T_P(V) \geq x$. We use $S_0(\cdot)$ to denote S_{P_0} from here on.

Define cutoff η_P as

$$\eta_P = S_P^{-1}(K - \underline{K_{A,P}})$$

The assumptions below in Sect. 1.5 guarantee that $S_P^{-1}(K - \underline{K_{A,P}})$ exists and η_P is well defined. η is set so that there is a total cost K of treating with $Z = 1$ everyone having $T(V) \geq \eta$. Further let:

$$\tau_P = \max\{\eta_P, 0\}$$

Thus, τ gives the cutoff for the scoring function $T(V)$, so the optimal rule is

$$d_P(V) = 1 \text{ iff } T_P(V) \geq \tau_P$$

Lemma 2 *Assume (A2)–(A4). Then, the optimal decision rule d_0 for parameter $\Psi_{Z,0}$ as defined in Eq. 1.1 is the deterministic solution $d_0(V) = 1$ iff $T_0(V) \geq \tau_0$, with T_0, τ_0 as defined above.*

The proof is given in Toth (2016). That work also describes modifications to the optimal solution for d_0 when Z is continuous or categorical.

1.4.1 Closed-Form Solution for Optimal Treatment Rule $A = d_0(V)$

The solution given above goes through for the case of intervening on the treatment, with the two main modifications that: (1) replace intervention variable Z with A, and (2) replace $\mu_b(W)$ with $m(W)$. These latter quantities represent the effect on Y of applying the intervention versus the baseline treatment (at Z or A, respectively).

1.5 Assumptions for Pathwise Differentiability of $\Psi_{Z,0}$ and $\Psi_{A,0}$

We use notation $d_0 = d_{P_0}$, $\tau_0 = \tau_{P_0}$, etc. We state these assumptions for $\Psi_{Z,0}$. The exact same assumptions apply for $\Psi_{A,0}$, replacing Z with A in a few places.

These three assumptions are needed to ensure pathwise differentiability and prove the form of the canonical gradient (Theorem 1).

Assumptions (A2)–(A4).

(***A2***) Positivity assumption: $0 < \rho_0(W) < 1$.

(***A3***) There is a neighborhood of η_0 where $S_0(x)$ is Lipschitz continuous, and a neighborhood of $S_0(\eta_0) = K - \underline{K_Z}_0$ where $S_0^{-1}(y)$ is Lipschitz continuous.

(***A4***) $Pr_0(T_0(V) = \tau) = 0$ for all τ in a neighborhood of τ_0.

Note that (A3) implies that $S_0^{-1}(K - \underline{K_Z}_0)$ exists. Note also that (A3) actually implies $Pr_0(T_0(V) = \eta) = 0$ for η in a neighborhood of η_0, and thus, (A3) implies (A4) when $\eta_0 > 0$ and $\tau_0 = \eta_0$.

Need for (A4) (Guarantee of non-exceptional law).
If (A4) does not hold and there is positive probability of individuals being at the threshold for being treated or not under the optimal rule, then the solution $d(V)$ is not unique, and $\Psi_{Z,0}$ is no longer pathwise differentiable. It is easy to see that under (A4), the optimal $d(V)$ over the broader set of non-deterministic decision rules is a deterministic rule. Toth (2016) describes why (A4) is a reasonable assumption in practice when we have a constraint $\underline{K_Z} < K < \overline{K_Z}$ that allows for only a strict subset of the population to be treated.

1.6 TMLE for Optimal Intent-to-Treat Problem ($\Psi_{Z,0}$)

All proofs and derivations for what follows are given in Toth (2016).

1.6.1 Canonical Gradient for $\Psi_{Z,0}$

For $O = (W, Z, A, Y)$, and deterministic rule $d(V)$, define

$$D_1(d, P)(O) \triangleq \frac{I(Z = d(V))}{\rho_P(W)}(Y - \mu_P(Z, W)) \tag{1.4}$$

$$D_2(d, P)(O) \triangleq \mu_P(d(V), W) - E_P\mu_P(d(V), W) \tag{1.5}$$

$$D_3(d, \tau, P)(O) = -\tau(c_T(d(V), W) - K) \tag{1.6}$$

Define

$$D^*(d, \tau, P)(O) \triangleq D_1(d, P)(O) + D_2(d, P)(O) + D_3(d, \tau, P)(O)$$

Theorem 1 *Assume (A1)–(A4) above. Then Ψ_Z is pathwise differentiable at P_0 with canonical gradient $D_0 = D^*(d_0, \tau_0, P_0)$.*

1.6.2 TMLE

The relevant components for estimating $\Psi_Z = E_W\mu(Z = d(V), W)$ are $Q = (P_W, \mu(Z, W))$. Decision rule d is also part of Ψ, but it is a function of P_W, $\mu(Z, W)$. The nuisance parameter is $g = \rho(W)$. First convert Y to the unit interval via a linear transformation $Y \to \tilde{Y}$, so that $\tilde{Y} = 0$ corresponds to $Y_{\min}$ and $\tilde{Y} = 1$ to $Y_{\max}$. We assume $Y \in [0, 1]$ from here.

1. Use the empirical distribution $P_{W,n}$ to estimate P_W. Make initial estimates of $\mu_n(Z, W)$ and $g_n = \rho_n(W)$ using any strategy desired. Data-adaptive learning using Super Learner is recommended.
2. The empirical estimate $P_{W,n}$ gives an estimate of $Pr_{V,n}(V) = E_{W,n}I(F_V(W) = V)$, $\underline{K_{Z,n}} = E_{W,n}c_T(0, W)$, $\overline{K_{Z,n}} = E_{W,n}c_T(1, W)$, and $c_{b,Z,n}(V) = E_{W,n|V}(c_T(\overline{1, W}) - c_T(0, W))$.
3. Estimate $\mu_{b,0}$ as $\mu_{b,n}(V) = E_{W,n|V}(\mu_n(1, W) - \mu_n(0, W))$.
4. Estimate $T_0(V)$ as $T_n(V) = \frac{\mu_{b,n}(V)}{c_{b,Z,n}(V)}$.
5. Estimate $S_0(x)$ using $S_n(x) = E_{V,n}[I(T_n(V) \geq x)(c_{b,Z,n}(V)]$.
6. Estimate η_0 as η_n using $\eta_n = S_n^{-1}(K - \underline{K_{Z,n}})$ and $\tau_n = \max\{0, \eta_n\}$.

7. Estimate the decision rule as $d_n(V) = 1$ iff $T_n(V) \geq \tau_n$.
8. Now fluctuate the initial estimate of $\mu_n(Z, W)$ as follows: For $Z \in [0, 1]$, define covariate $H(Z, W) \triangleq \frac{I(d_n(V)=Z)}{g_n(W)}$. Run a logistic regression using:

   Outcome: $(Y_i : i = 1, \ldots, n)$
   Offset: $(\text{logit}\, \mu_n(Z_i, W_i), i = 1, \ldots, n)$
   Covariate: $(H(Z_i, W_i) : i = 1, \ldots, n)$

   Let ε_n represent the level of fluctuation, with $\varepsilon_n = \text{argmax}_\varepsilon \frac{1}{n} \sum_{i=1}^{n} [\mu_n(\varepsilon)(Z_i, W_i) \log Y_i + (1 - \mu_n(\varepsilon)(Z_i, W_i)) \log(1 - Y_i)]$ and $\mu_n(\varepsilon)(Z, W) = \text{logit}^{-1}(\text{logit}\, \mu_n(Z, W) + \varepsilon H(Z, W))$.
9. Set the final estimate of $\mu(Z, W)$ to $\mu_n^*(Z, W) = \mu_n(\varepsilon_n)(Z, W)$.
10. Finally, form final estimate of $\Psi_{Z,0} = \Psi_{Z,d_0}(P_0)$ using the plug-in estimator

$$\Psi_Z^* = \Psi_{Z,d_n}(P_n^*) = \frac{1}{n} \sum_{i=1}^{n} \mu_n^*(Z = d_n(V_i), W_i)$$

We have used the notation $\Psi_{Z,d}(P)$ referring to mean outcome under decision rule $Z = d(V)$, and Ψ_n^* the final estimate of the data-generating distribution.

It is easy to see that $P_n D^*(d_n, \tau_n, P_n^*) = 0$: We have $P_n D_1(d_n, P_n^*) = P_n \frac{d}{d\varepsilon} L(Q_n(\varepsilon|g_n), g_n, (O_1, \ldots, O_n))|_{\varepsilon=0} = 0$; $P_n D_2(d_n, P_n^*) = 0$ when we are using the empirical distribution $P_{W,n}$; and $P_n D_3(d_n, \tau_n, P_n^*) = 0$ is described in the proof of optimality of the closed-form solution in Toth (2016).

1.6.3 Theoretical Results for Ψ_Z^*

1.6.3.1 Conditions for Efficiency of Ψ_Z^*

These six conditions are needed to prove asymptotic efficiency (Theorem 2). As discussed in Toth (2016), when all relevant components and nuisance parameters ($P_{W,n}$, ρ_n, μ_n) are consistent, then (C3) and (C4) hold, while (C6) holds by construction of the TMLE estimator.

(C1) $\rho_0(W)$ satisfies the strong positivity assumption: $Pr_0(\delta < \rho_0(W) < 1 - \delta) = 1$ for some $\delta > 0$.

(C2) The estimate $\rho_n(W)$ satisfies the strong positivity assumption, for a fixed $\delta > 0$ with probability approaching 1, so we have $Pr_0(\delta < \rho_n(W) < 1 - \delta) \to 1$.

Define second-order terms as follows:

$$R_1(d, P) \triangleq E_{P_0}\left[\left(1 - \frac{Pr_{P_0}(Z = d|W)}{Pr_P(Z = d|W)}\right)(\mu_P(Z = d, W) - \mu_0(Z = d, W))\right]$$

$$R_2(d, \tau_0, P) \triangleq E_{P_0}\Big[(d - d_0)(\mu_{b,0}(V) - \tau_0 c_{b,0}(V))\Big]$$

Let $R_0(d, \tau_0, P) = R_1(d, P) + R_2(d, \tau_0, P)$.

(C3) $R_0(d_n, \tau_0, P_n^*) = o_{P_0}(n^{-\frac{1}{2}})$.

(C4) $P_0[(D^*(d_n, \tau_0, P_n^*) - D_0)^2] = o_{P_0}(1)$.

C5) $D^*(d_n, \tau_0, P_n^*)$ belongs to a P_0-Donsker class with probability approaching 1.

(C6) $\frac{1}{n}\sum_{i=1}^n D^*(d_n, \tau_0, P_n^*)(O_i) = o_{P_0}(n^{-\frac{1}{2}})$.

1.6.3.2 Sufficient Conditions for Lemma 3

(E1) GC-like property for $c_{b,Z}(V)$, $\mu_{b,n}(V)$:
$\sup_V |(E_{W,n|V} - E_{W,0|V})c_{b,T}(W)| = \sup_V(|c_{b,Z,n}(V) - c_{b,Z,0}(V)|) = o_{P_0}(1)$

(E2) $\sup_V |E_{W,0|V}\mu_{b,n}(W) - E_{W,0|V}\mu_{b,0}(W)| = o_{P_0}(1)$

(E3) $S_n(x)$, defined as $x \to E_{V,n}[I(T_n(V) \geq x)c_{b,Z,n}(V)]$ is a GC-class.

(E4) Convergence of ρ_n, μ_n to ρ_0, μ_0, respectively, in $L^2(P_0)$ norm at a $O(n^{-1/2})$ rate in each case.

When all relevant components and nuisance parameters are consistent, as is the case when Theorem 2 below holds and our estimator is efficient, we also expect conditions (E1)–(E4) to hold.

Toth (2016) discusses the assumptions and conditions above in detail.

1.6.3.3 Efficiency and Inference

Theorem 2 (Ψ_Z^* is asymptotically linear and efficient.) *Assume assumptions (A1)–(A4) and conditions (C1)–(C6). Then, $\Psi_Z^* = \Psi_Z(P_n^*) = \Psi_{Z,d_n}(P_n^*)$ as defined by the TMLE procedure is a RAL estimator of $\Psi_Z(P_0)$ with influence curve D_0, so*

$$\Psi_Z(P_n^*) - \Psi_Z(P_0) = \frac{1}{n}\sum_{i=1}^n D_0(O_i) + o_{P_0}(n^{-\frac{1}{2}}).$$

Further, Ψ_Z^* is efficient among all RAL estimators of $\Psi_Z(P_0)$.

Inference. Let $\sigma_0^2 = Var_{O\sim P_0} D_0(O)$. By Theorem 2 and the central limit theorem, $\sqrt{n}(\Psi_Z(P_n^*) - \Psi_Z(P_0))$ converges in distribution to a $N(0, \sigma_0^2)$ distribution. Let $\sigma_n^2 = \frac{1}{n}\sum_{i=1}^n D^*(d_n, \tau_n, P_n^*)(O_i)^2$ be an estimate of σ_0^2.

Lemma 3 *Under the assumptions (C1) and (C2), and conditions (E1)–(E4), we have* $\sigma_n \to_{P_0} \sigma_0$. *Thus, an asymptotically valid 2-sided* $1-\alpha$ *confidence interval is given by*

$$\Psi_Z^* \pm z_{1-\frac{\alpha}{2}} \frac{\sigma_n}{\sqrt{n}}$$

where $z_{1-\frac{\alpha}{2}}$ denotes the $(1-\frac{\alpha}{2})$-quantile of a $N(0,1)$ r.v.

1.6.3.4 Double Robustness of $\Psi^*_{Z,n}$

Theorem 2 demonstrates consistency and efficiency when all relevant components and nuisance parameters are consistently estimated. Another important issue is under what cases of partial misspecification we still get a consistent estimate of $\Psi_{Z,0}$, albeit an inefficient one. Our TMLE-based estimate Ψ_Z^* is a consistent estimate of $\Psi_{Z,0}$ under misspecification of $\rho_n(W)$ in the initial estimates, but not under misspecification of $\mu_n(W, Z)$. However, it turns out there is still an important double robustness property. If we consider $\Psi_Z^* = \Psi_{Z,d_n}(P_n^*)$ as an estimate of $\Psi_{Z,d_n}(P_0)$, where the optimal decision rule $d_n(V)$ is estimated from the data, then we have that Ψ_Z^* is double robust to misspecification of ρ_n or μ_n in the initial estimates.

Lemma 4 (Ψ_Z^* is a double robust estimator of $\Psi_{Z,d_n}(P_0)$.) *Assume assumptions (A1)–(A4) and conditions (C1)–(C2). Also assume the following version of (C4):* $Var_{O\sim P_0}(D_1(d_n, P_n^*)(O) + D_2(d_n, P_n^*)(O)) < \infty$.
Then, $\Psi_Z^* = \Psi_{Z,d_n}(P_n^*)$ *is a consistent estimator of* $\Psi_{Z,d_n}(P_0)$ *when either* μ_n *is specified correctly, or* ρ_n *is specified correctly.*

The proof of this lemma is based on the equation

$$\Psi_{Z,d_n}(P_n^*) - \Psi_{Z,d_n}(P_0) = -P_0\big[D_1(d_n, P_n^*) + D_2(d_n, P_n^*)\big] + R_1(d_n, P_n^*)$$

where D_1, D_2, and R_1 are as defined in Sects. 1.6.1 and 1.6.3.1.

1.7 TMLE for Optimal Treatment Problem ($\Psi_{A,0}$)

We now present results for the case of intervening on the treatment, setting $A = d(V)$.

1.7.1 Efficient Influence Curve $D_A^*(\Psi_0)$

Lemma 5 *Let*

$$J_0(Z, W) = \frac{I(Z=1)}{\rho_0(W)} + \frac{\left(\frac{I(Z=1)}{\rho_0(W)} - \frac{I(Z=0)}{1-\rho_0(W)}\right)\left(d_0(V) - \Pi_0(W, Z=1)\right)}{\Pi_0(W, Z=1) - \Pi_0(W, Z=0)}$$

The efficient influence curve $D_A^*(\Psi_0)$ *is*

$$D_A^*(\Psi_0) = -\tau_0 E_{P_0}[c_T(d_0(V), W) - K] \tag{1.7}$$
$$+m_0(W)d_0(V) + \theta_0(W) - \Psi_0 \tag{1.8}$$
$$-J_0(Z, W)m_0(W)\left[A - \Pi_0(W, Z)\right] \tag{1.9}$$
$$+J_0(Z, W)\left[Y - (m_0(W)\Pi_0(W, Z) - \theta_0(W))\right] \tag{1.10}$$

We also write $D^*(d_0, \tau_0, P_0)$. For convenience, denote lines (1)–(4) of D^* above as D_c^*, D_W^*, D_Π^*, and D_μ^*, respectively. Finally, let D_{A,d_n}^* denote the efficient influence curve for $\Psi_{A,d_n}(P_0) \triangleq E_{P_{W,0}} m_0(W)d_n(V) + \theta_0(W)$, which is the mean counterfactual estimate when the decision rule is estimated from the data. We have $D_{A,d_n}^* = D_W^* + D_\Pi^* + D_\mu^*$ (see Toth 2016).

1.7.2 Iterative TMLE Estimator

We have derived two different TMLE-based estimators for $\Psi_{A,0}$. We present an iterative estimator here, which involves a standard, numerically well-behaved, and easily understood likelihood maximization operation at each step. The other estimator uses a logistic fluctuation in a single non-iterative step and has the advantage that the estimate μ respects the bounds of Y found in the data (see Toth 2016; Toth and van der Laan 2016).

The relevant components for estimating $\Psi_A = E_W[m(W)d(V) + \theta(W)]$ are $Q = (P_W, m, \theta)$. The nuisance parameters are $g = (\rho, \Pi)$. $d(V)$ and τ can be thought of as functions of P_W, m here. Let

$h_1(W) \triangleq \frac{1}{\rho(W)(\Pi(W,1)-\Pi(W,0))} + \frac{d(V)-\Pi(W,1)}{(\Pi(W,1)-\Pi(W,0))^2}\frac{1}{\rho(W)(1-\rho(W))}$. Also, let $h_2(W) \triangleq \frac{1}{\rho}\left[1 - \frac{\Pi(W,1)}{\Pi(W,1)-\Pi(W,0)} + \frac{d-\Pi(W,1)}{\Pi(W,1)-\Pi(W,0)}(1 - \frac{\Pi(W,1)}{\Pi(W,1)-\Pi(W,0)}\frac{1}{1-\rho})\right]$.
Then, we have that $D_\mu^* = (h_1\Pi + h_2)(Y - m\Pi - \theta)$.

If A is not binary, convert A to the unit interval via a linear transformation $A \to \tilde{A}$ so that $\tilde{A} = 0$ corresponds to $A_{\min}$ and $\tilde{A} = 1$ to $A_{\max}$. We assume $A \in [0, 1]$ from here.

1. Use the empirical distribution $P_{W,n}$ to estimate P_W. Make initial estimates of $Q = \{m_n(W), \theta_n(W)\}$ and $g_n = \{\rho_n(W), \Pi_n(W, Z)\}$ using any strategy desired. Data-adaptive learning using Super Learner is recommended.

2. The empirical estimate $P_{W,n}$ gives an estimate of $Pr_{V,n}(V) = E_{W,n}I(F_V(W) = V)$, $K_{A,n} = E_{W,n}c_A(0, W)$, $\overline{K_{A,n}} = E_{W,n}c_A(1, W)$, and $c_{b,A,n}(V) = E_{W,n|V}(c_A(\overline{1, W}) - c_A(0, W))$.
3. Estimate $m_n(V)$ as $E_{W,n|V}m(W)$.
4. Estimate $T_0(V)$ as $T_n(V) = \frac{m_n(V)}{c_{b,A,n}(V)}$.
5. Estimate $S_0(x)$ using $S_n(x) = E_{V,n}[I(T_n(V) \geq x)(c_{b,A,n}(V))]$.
6. Estimate η_0 as using $\eta_n = S_n^{-1}(K - K_{A,n})$ and $\tau_n = \max\{0, \eta_n\}$.
7. Estimate the decision rule as $d_n(V) = \overline{1}$ iff $T_n(V) \geq \tau_n$ (the decision rule is not updated iteratively).

ITERATE STEPS (8)–(9) UNTIL CONVERGENCE:

8. Fluctuate the initial estimate of $m_n(W)$, $\theta_n(W)$ as follows: Using $\mu_n(Z, W) = m_n(W)\Pi_n(Z, W) + \theta_n(W)$, run an OLS regression:
   Outcome: $(Y_i : i = 1, \ldots, n)$
   Offset: $(\mu_n(Z_i, W_i), i = 1, \ldots, n)$
   Covariate: $(h_1(W_i)\Pi_n(Z_i, W_i) + h_2(W_i) : i = 1, \ldots, n)$
   Let ε_n represent the level of fluctuation, with
   $\varepsilon_n = \text{argmax}_\varepsilon \frac{1}{n}\sum_{i=1}^n (Y_i - \mu_n(\varepsilon)(Z_i, W_i))^2$ and
   $\mu_n(\varepsilon)(Z, W) = \mu_n(Z, W) + \varepsilon(h_1(W)\Pi_n(Z, W) + h_2(W))$.

   Note that $\mu_n(\varepsilon) = (m_n + \varepsilon h_1)\Pi_n + (\theta_n + \varepsilon h_2)$ stays in the semiparametric regression model.
   Update m_n to $m_n(\varepsilon) = m_n + \varepsilon h_1$, θ_n to $\theta_n(\varepsilon) = \theta_n + \varepsilon h_2$.
9. Now fluctuate the initial estimate of $\Pi_n(Z, W)$ as follows: Use covariate $J(Z, W)$ as defined in Lemma 5. Run a logistic regression using:

   Outcome: $(A_i : i = 1, \ldots, n)$
   Offset: $(\text{logit}\, \Pi_n(Z_i, W_i), i = 1, \ldots, n)$
   Covariate: $(J(Z_i, W_i)m(W_i) : i = 1, \ldots, n)$
   Let ε_n represent the level of fluctuation, with
   $\varepsilon_n = \text{argmax}_\varepsilon \frac{1}{n}\sum_{i=1}^n [\Pi_n(\varepsilon)(Z_i, W_i) \log A_i + (1 - \Pi_n(\varepsilon)(Z_i, W_i))$
   $\log(1 - A_i)]$ and $\Pi_n(\varepsilon)(Z, W) = \text{logit}^{-1}(\text{logit}\, \Pi_n(Z, W) + \varepsilon J(Z, W)m(W))$.
   Update Π_n to $\Pi_n(\varepsilon)$. Also update $h_1(W)$, $h_2(W)$ to reflect the new Π_n.

---

10. Finally, form final estimate of $\Psi_{A,0} = \Psi_{A,d_0}(P_0)$, using a plug-in estimator with the final estimates upon convergence m_n^* and θ_n^*:

$$\Psi_A^* = \Psi_{A,d_n}(P_n^*) = \frac{1}{n}\sum_{i=1}^n \left[m_n^*(W_i) \cdot d_n(V_i) + \theta_n^*(W_i)\right]$$

As for Ψ_Z, it is straightforward to check that the efficient influence equation $P_n D^*(d_n, \tau_n, P_n^*) = 0$.

1.7.3 Double Robustness of Ψ_A^*

As in Sect. 1.6.3.4, Ψ_A^* is not a double robust estimator of $\Psi_{A,0}$: Component $m(W)$ must always be consistently specified as a necessary condition for consistency of Ψ_A^*. However, if we consider $\Psi_A^* = \Psi_{A,d_n}(P_n^*)$ as an estimate of $\Psi_{A,d_n}(P_0)$, where the optimal decision rule $d_n(V)$ is estimated from the data, then we have that Ψ_A^* is double robust:

Lemma 6 (Ψ_A^* is a double robust estimator of $\Psi_{A,d_n}(P_0)$.) *Assume (A1)–(A4) and (C1)–(C2). Also assume* $Var_{O\sim P_0}(D_d^*(d_n, P_n^*)(O)) < \infty$.
Then, $\Psi_A^* = \Psi_{A,d_n}(P_n^*)$ *is a consistent estimator of* $\Psi_{A,d_n}(P_0)$ *when either:*

- m_n *and* θ_n *are consistent*
- ρ_n *and* Π_n *are consistent*
- m_n *and* ρ_n *are consistent*

Above D_d^* refers to $D_\mu^* + D_\Pi^* + D_W^*$, the portions of the efficient influence curve that are orthogonal to variation in decision rule d. The proof is straightforward (see Toth 2016).

1.8 Simulations

1.8.1 Setup

We use two main data-generating functions:

Dataset 1 (categorical Y).

Data is generated according to:

$$
\begin{aligned}
U_{AY} &\sim \text{Bernoulli}(1/2)\\
W1 &\sim \text{Uniform}(-1, 1)\\
W2 &\sim \text{Bernoulli}(1/2)\\
Z &\sim \text{Bernoulli}(\alpha)\\
A &\sim \text{Bernoulli}(W1 + 10 \cdot Z + 2 \cdot U_{AY} - 10)\\
Y &\sim \text{Bernoulli}((1 - A) * (\text{plogis}(W2 - 2 - U_{A,Y})) + (A) * (\text{plogis}(W1 + 4))
\end{aligned}
$$

$U_{A,Y}$ is the confounding term. For the simulations where $V \subset W$, we take $V = (1(W1 \geq 0) + -1(W1 < 0), W2)$. We have $c_T(Z = 1, W) = 1$, $c_T(Z = 0, W) = 0$ for all W here.

Dataset 2 (continuous Y.)

We use three-dimensional W and distribution

$$\begin{aligned} U_{AY} &\sim \text{Normal}(0, 1) \\ W &\sim \text{Normal}(\mu_\beta, \Sigma) \\ Z &\sim \text{Bernoulli}(0.1) \\ A &\sim -2 \cdot W1 + W2^2 + 4 \cdot W3 \cdot Z + U_{AY} \\ Y &\sim 0.5 \cdot W1 \cdot W2 - W3 + 3 \cdot A \cdot W2 + U_{AY} \end{aligned}$$

When $V \subset W$, we use either V equals $W1$ rounded to the nearest 0.2, or alternately, V is $W3$ rounded to the nearest 0.2. We also have $c_T(0, W) = 0$ for all W, and $c_T(1, W) = 1 + b \cdot W1$, and varying μ_β, Σ, and b.

Forming initial estimates.
We use the empirical distribution $P_{W,n}$ for the distribution of W. For learning μ_n, we use Super Learner, with the following libraries of learners (the names of learners are as specified in the SuperLearner package (van der Laan et al. 2007):

For continuous Y: glm, step, randomForest, nnet, svm, polymars, rpart, ridge, glmnet, gam, bayesglm, loess, mean.

For categorical Y: glm, step, svm, step.interaction, glm.interaction, nnet.4, gam, randomForest, knn, mean, glmnet, rpart.

Further, we included different parameterizations of some of the learners given above, such as ntree = 100, 300, 500, 1000 for randomForest.

Finally, for learning ρ_n, we use a correctly specified logistic regression, regressing Z on W (except for simulation (C) as described below).

Estimators used.
For both parameters of interest Ψ_Z and Ψ_A, we report results on the TMLE estimator Ψ_Z^* (or Ψ_A^*), and the initial substitution estimator $\Psi_{Z,n}^0$ (or $\Psi_{A,n}^0$). The latter is the plug-in estimate, for instance $\Psi_{Z,n}^0 \triangleq \Psi_Z(P_{W,n}, \mu_n)$, that uses the same initial estimates of relevant components and the nuisance parameter as TMLE. Thus, the initial substitution estimator gives a comparison of TMLE to a straightforward semiparametric, machine learning-based approach. 1000 repetitions are done of each simulation.

Table 1.1 (Simulation A.) Consistent estimation of $\Psi_{Z,0}$ using machine learning, categorical $Y \cdot \Psi_{Z,0} = 0.3456$, $K = 0.3$, and $V \subset W \cdot \sigma_n^2 = \mathrm{Var}_{O\sim P_n} D_Z^*(d_n, \tau_n, P_n^*)(O)$

N = 250					
Estimator	Ψ_Z^*	Bias	Var	σ_n^2/N	Cover
TMLE	0.3545	0.0089	0.0071	0.0010	88.3
CV-TMLE	0.3541	0.0085	0.0017	0.0010	90.6
Init. Substit.	0.3427	−0.0029	0.0067	0.0010	(87.9)
N = 1000					
TMLE	0.3485	0.0029	0.0003	0.0003	93.3
CV-TMLE	0.3497	0.0041	0.0002	0.0003	96.8
Init. Substit.	0.3344	−0.0112	0.0003	0.0003	(88.3)
N = 4000					
TMLE	0.3467	0.0011	0.0001	0.0001	95.0
CV-TMLE	0.3498	0.0002	0.0001	0.0001	94.7
Init. Substit.	0.3429	−0.0027	0.0001	0.0001	(93.3)

Table 1.2 (Simulation B.) Consistent estimation of $\Psi_{A,0}$ using machine learning, continuous $Y \cdot \Psi_{A,0} = 336.2$, $K = 0.8$, and $V \subset W \cdot \sigma_n^2 = \mathrm{Var}_{O\sim P_n} D_Z^*(d_n, \tau_n, P_n^*)(O)$

N = 250					
Estimator	Ψ_A^*	Bias	Var	σ_n^2/N	Cover
TMLE	327.5	−8.7	344.7	176.3	78.4
Init. Substit.	310.0	−26.2	495.1	174.0	(47.8)
N = 1000					
TMLE	332.9	−3.3	40.7	38.5	89.0
Init. Substit.	322.7	−13.5	126.8	43.1	(53.2)
N = 4000					
TMLE	334.5	−1.7	8.4	9.1	93.3
Init. Substit.	328.7	−7.5	25.9	8.8	(41.3)

Simulations (A–B): using a large library of learning algorithms for consistent initial estimates.

Tables 1.1 and 1.2 show the behavior of our estimators when machine learning is used to consistently estimate all relevant components and nuisance parameters. Table 1.1 deals with estimating Ψ_Z when Y is categorical. In this case, bias is very low with or without the TMLE fluctuation step. σ_n^2/n gives a consistent estimate of the variance of Ψ_Z^*, in this case where efficiency holds. We see that both estimators have very low variance that converges to σ_n^2/n by $n = 1000$. Coverage of 95% confidence intervals is also displayed, with intervals calculated as $\Psi_n^* \pm 1.96\frac{\sigma_n}{\sqrt{n}}$, as in Lemma 3. The coverage is given in parentheses for the initial substitution estimator, as σ_n^2 is not necessarily the right variance. The TMLE estimators show better coverage,

even though, in this example, the width of the confidence intervals was accurate for all estimators for $n \geq 1000$. This may be due to the asymptotic linearity property of the TMLE-based estimators, ensuring that they follow a normal distribution as n becomes large.

Y is continuous in Table 1.2. TMLE convincingly outperforms the initial substitution estimator in both bias and variance here. Only the TMLE estimator is guaranteed to be efficient, and we see a significant improvement in variance. The estimated asymptotic variance σ_n^2/n approximates the variance seen in Ψ_A^* fairly well for $n \geq 1000$. The coverage of confidence intervals for TMLE seems to converge to 95% more slowly than for the previous case of categorical Y.

Simulation (C): double robustness under partial misspecification.
As described in Sect. 1.7.3, $\Psi_A^* = \Psi_{A,d_n}^*$ is a double robust estimator of $\Psi_{A,d_n}(\Psi_0)$, but not necessarily of $\Psi_{A,0}$.

Table 1.3 verifies consistency of Ψ_A^* when the initial estimate for μ_n is grossly misspecified as $\mu_n = \text{mean}(Y)$. This creates a discrepancy of $\sim$0.1 points between $\Psi_{A,d_n}(P_0)$ and $\Psi_{A,0}$. The initial substitution estimator retains a bias of around -0.09 in estimating $\Psi_{A,d_n}(P_0)$, while TMLE demonstrates practically zero bias by $n = 1000$. TMLE is not efficient in this setting of partial misspecification. It has significantly larger variance than the initial substitution estimator for smaller sample sizes, but the variances are similar by $n = 4000$. For confidence intervals, the width was calculated by estimating $\text{Var}(\Psi_{A,d_n})$ as $\sigma_n^2 = \text{Var}_{O\sim P_n} D_{d_n}^*(P_n^*)(O)$, where $D_{d_n}^*(P)$ is the effi-

Table 1.3 (Simulation C.) Robustness of Ψ_A^* to partial misspecification, μ_n is misspecified. $\Psi_{A,0} = 0.63$, $K = 0.5$, and $V = W$

N = 1000					
Estimator	Ψ_A^*	$\left(\Psi^* - \Psi_{d_n}(P_0)\right)$	$\left(\Psi^* - \Psi_0\right)$	Var	Cover
TMLE	0.54	0.00	−0.09	0.69	93.3
Init. Substit.	0.45	−0.10	−0.18	0.24	(69.2)
N = 4000					
TMLE	0.54	0.00	−0.09	0.11	96.8
Init. Substit.	0.45	−0.09	−0.18	0.10	(40.1)

Table 1.4 (Simulation D.) Estimation of true mean outcome $\Psi_{Z,d_n}(P_0)$, under rule $d_n \cdot \Psi_{Z,d_n}(P_0) = 162.8$ when $K = 0.2$, and $\Psi_{Z,d_n}(P_0) = 289.1$ when $K = 0.8$. Sample size is $N = 1000$ and $V = W$

	K = 0.2		K = 0.8	
Learning μ_n	Ψ_Z^*	Var	Ψ_Z^*	Var
Large library	158.9	8.14	286.4	9.32
Small library	148.3	49.45	267.9	16.28
No fitting	142.2	12.83	264.1	10.30

cient influence curve of $\Psi_{A,d_n}(P)$ as defined in Sect. 1.7.1. It provides a conservative (over)-estimate of variance for confidence intervals, as discussed in Toth and van der Laan (2016). We see that TMLE's coverage converges to just above 95%. On the other hand, coverage is very low for the initial substitution estimator due to its bias. This is despite the fact that the intervals are too wide in this case.

Simulation (D): quality of the estimate of d_n versus the true mean outcome attained under rule d_n.
We study how more accurate estimation of the decision rule d_n can lead to a higher objective obtained. The objective maximized here is the mean outcome under rule d_n, where d_n must satisfy a cost constraint. We use the known true distributions for $P_{W,0}$ and μ_0 in calculating the value of mean outcome under d_n as $\Psi_{d_n}(P_0) = E_{P_0}\mu_0(W, Z = d_n(V))$. The highest the true mean outcome can be under a decision rule that satisfies $E_{P_0}c_T(W, Z = d(V)) \leq K$ is Ψ_0 using optimal rule $d = d_0$. Therefore, the discrepancy between $\Psi_{d_n}(P_0)$ and Ψ_0 gives a measure of how inaccurate estimation of the decision rule diminishes the objective.

We compare $\Psi_{d_n}(P_0)$ when estimating μ_n using the usual large library of learners; when using a smaller library of learners consisting of mean, loess, nnet.size = 3, nnet.size = 4, nnet.size = 5; and finally when we set $\mu_n = \text{mean}(Y) \cdot d_n$ is estimated using μ_n as usual (note that it is the same between the initial substitution and TMLE-based estimates). Table 1.4 confirms the importance of forming a good fit with the data for achieving a high mean outcome. For $K = 0.2$ when roughly 20% of the population could be assigned $Z = 1$, the mean outcome was only a few points below the true optimal mean outcome Ψ_{d_0}, when using the full library of learners (158.9 vs. 162.8). However, it was about 15 points lower when using a much smaller library of learners. In fact, even when using machine learning with several nonparametric methods in the case of the smaller library, the objective $\Psi_{d_n}(P_0)$ attained was not far from that attained with the most uninformative $\mu_n = \text{mean}(Y)$. Very similar results hold for the less constrained case of $K = 0.8$.

1.9 Discussion

We considered the resource-allocation problem of finding the optimal mean counterfactual outcome given a general cost constraint, in the setting where unmeasured confounding is a possibility and an instrumental variable is available. This work dealt with both problems of finding an optimal treatment regime, and finding the optimal intent-to-treat regime. For both cases, we gave closed-form solutions of the optimal intervention and derived estimators for the optimal mean counterfactual outcome. Our model allows the individualized treatment (or intent-to-treat) rules to be a function of an arbitrary subset of baseline covariates. Estimation is done using the targeted maximum likelihood (TMLE) methodology, which is a semiparametric approach having a number of desirable properties (efficiency, robustness to misspecification, asymptotic normality, and being a substitution estimator). Simulation

results showed that TMLE can simultaneously demonstrate both finite-sample bias reduction and lower variance than straightforward machine learning approaches. The empirical variance of TMLE estimators appears to converge to the semiparametric efficiency bound, and confidence intervals are accurate for sample sizes of a few thousand. Consistency in the case of partial misspecification was confirmed, in the sense of Lemmas 4 and 6. Our simulations also addressed the important question of to what extent improved statistical estimation can lead to better optimization results. We were able to demonstrate significant increases in the value of the mean outcome under the estimated optimal rule, when a larger library of data-adaptive learners achieved a closer fit.

References

Angrist, J. D., & Krueger, A. B. (1991). Does compulsory school attendance affect schooling and earnings? *The Quarterly Journal of Economics*, *106*(4), 979–1014.

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, *91*, 444–471.

Brookhart, M. A., & Schneeweiss, S. (2007). Preference-based instrumental variable methods for the estimation of treatment effects. *International Journal of Biostatistics*, *3*(1), 1–14.

Brookhart, M. A., Rassen, J. A., & Schneeweiss, S. (2010). Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and Drug Safety*, *19*(6), 537–554.

Chakraborty, B., & Moodie, E. E. (2013). *Statistical methods for dynamic treatment regimes*. Berlin Heidelberg New York: Springer.

Chakraborty, B., Laber, E., & Zhao, Y. (2013). Inference for optimal dynamic treatment regimes using an adaptive m-out-of-n bootstrap scheme. *Biometrics*, *69*(3), 714–723.

Chesney, M. A. (2006). The elusive gold standard. Future perspectives for HIV adherence assessment and intervention. *Journal of Acquired Immune Deficiency Syndromes*, *43*(1), S149–155.

Editors: National Research Council (US) Committee on A Framework for Developing a New Taxonomy of Disease. (2011). Toward precision medicine: Building a knowledge network for biomedical research and a new taxonomy of disease. National Academies Press (US), Washington DC.

Gruber, S., & van der Laan, M. (2010). A targeted maximum likelihood estimator of a causal effect on a bounded continuous outcome. *International Journal of Biostatistics*, **6**(1). Article 26.

Imbens, G. W., & Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*, *62*, 467–475.

Karp, R. (1972). *Reducibility among combinatorial problems*. New York Berlin Heidelberg: Springer.

Luedtke, A., & van der Laan, M. (2016a). Optimal individualized treatments in resource-limited settings. *International Journal of Biostatistics*, *12*(1), 283–303.

Luedtke, A., & van der Laan, M. (2016b). Statistical inference for the mean outcome under a possibly non-unique optimal treatment strategy. *Annals of Statistics*, *44*(2), 713–742.

Newey, W. (1990). Semiparametric efficiency bounds. *Journal of Applied Econometrics*, *5*(2), 99–135.

Pearl, J. (2000). Causality: Models, reasoning, and inference. Cambridge University Press.

Robins, J. M. (2004). Optimal structural nested models for optimal sequential decisions. in *Proceeding of the Second Seattle Symposium in Biostatistics* (Vol. 179, pp. 189–326).

Toth, B. (2016). Targeted learning of individual effects and individualized treatments using an instrumental variable. PhD dissertation, U.C. Berkeley.

Toth, B., & van der Laan, M. (2016). TMLE for marginal structural models based on an instrument. U.C. Berkeley Division of Biostatistics Working Papers Series, working paper 350.

van der Laan, M., & Rose, S. (2011). *Targeted learning: Causal inference for observational and experimental data*. New York: Springer.

van der Laan, M., Rubin, D. (2006). Targeted maximum likelihood learning. *International Journal of Biostatistics*, *2*(1). Article 11.

van der Laan, M., Polley, E. C. & Hubbard, A. (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology*, *6*(1). Article 25.

Zhang, B., Tsiatis, A., Davidian, M., Zhang, M., & Laber, E. (2012). A robust method for estimating optimal treatment regimes. *Biometrics*, *68*, 1010–1018.