Behaviormetrika
https://doi.org/10.1007/s41237-021-00140-0

ORIGINAL PAPER

# Estimating the optimal treatment regime for student success programs

**Morten C. Wilke**[1,2] · **Richard A. Levine**[1] · **Maureen A. Guarcello**[1] · **Juanjuan Fan**[1]

Received: 27 October 2020 / Accepted: 1 July 2021


## Abstract

We expand methods for estimating an optimal treatment regime (OTR) from the personalized medicine literature to educational data mining applications. As part of this development, we detail and modify the current state-of-the-art, assess the efficacy of the approaches for student success studies, and provide practitioners the machinery to apply the methods in their specific problems. Our particular interest is to estimate an optimal treatment regime for students enrolled in an introductory statistics course at San Diego State University (SDSU). The available treatments are combinations of three programs SDSU implemented to foster student success in this large enrollment, bottleneck STEM course. We leverage tree-based reinforcement learning approaches based on either an inverse probability-weighted purity measure or an augmented probability-weighted purity measure. The thereby deduced OTR promises to significantly increase the average grade in the introductory course and also reveals the need for program recommendations to students as only very few, on their own, selected their optimal treatment.

**Keywords** Personalized learning · Reinforcement learning · Decision trees · Random forest · Supplemental instruction · Educational data mining

Communicated by Ryan Baker.

✉ Richard A. Levine
rlevine@sdsu.edu

Morten C. Wilke
morten.wilke@uni-ulm.de

[1] San Diego State University, San Diego, CA, USA

[2] Ulm University, Ulm, Baden-Württemberg, Germany

Published online: 14 July 2021

## 1 Introduction

In recent years, the strongly increasing availability of data on students made it possible for universities to assess their programs to aid student success through data-informed advising. One important field of assessment is for instance the *estimation of treatment effects* to study the efficacy of implemented programs. To this end, various methods have been deployed. We will mention just two that set up the methodological underpinnings for this paper. Spoon et al. (2016) leverages *random forests* to estimate the impact of supplemental instruction on student success in introductory statistics courses at San Diego State University (SDSU). Another widely applied method is *propensity score matching* as Powell et al. (2020) demonstrates to estimate the effect of changing the curriculum for a math course on student success.

In this paper, we bring—to our knowledge for the first time—the concept of *optimal treatment regimes* (OTR) from the personalized medicine literature to the educational data mining community to take the analysis from *assessing* programs to *recommending* programs. Loosely speaking, an OTR is a decision rule which recommends for every individual a treatment such that the mean potential outcome of the population is maximized. With respect to student success programs, our application considers performance in a specific course with the treatment regime including subject-specific one-on-one or small group tutoring, a Supplemental Instruction program, and an optional course-specific recitation class. A recommendation is relevant from a student's perspective, as students—especially those who are new to university—might struggle to decide which suite of student success programs is optimal for them. A recommendation is important from the university's perspective as it facilitates the efficient allocation of resources to those who need the programs and thereby potentially raising the success rate in this specific course. We will find such an OTR for students attending a bottleneck introductory statistics course at San Diego State University.

In recent years, various methods to estimate OTRs have been explored in the personalized medicine literature. To make our review of these approaches more comprehensible at this point in the paper, we consider the following example. A university offers a small number of tutoring hours per week to students enrolled in an introductory math course. To ensure that those tutoring hours are allocated to the students who need them most, the university conducts an analysis of data collected for the same math course in the previous semesters. The data include grade point average (GPA), age, tutoring attendance indicator, and the final course grade for every student in the math course. In the context of treatment regimes, there are two treatments available for every student in this example: whether the student shall attend tutoring or not.

Of importance, all following OTR approaches rely inherently on estimating the subgroup effects of the available treatments *and* interaction effects between students' characteristics and treatments on the grade via an outcome model such as a linear regression. If not, we would simply recommend for all students the treatment with the highest subgroup effect as optimal.

An intuitive first approach to identify optimal treatment regimes in our example is proposed by Qian and Murphy (2011): we predict the course grade with a regression model on the student characteristics, e.g., GPA and age, the chosen treatment, and interactions between characteristics and treatment. For a new student, the approach predicts a course grade for every treatment and estimates the OTR for this student as the one with the higher predicted course grade. To avoid overfitting, Qian and Murphy (2011) leverages a $l_1$-penalized regression. One drawback of this method, as Zhao et al. (2012) points out, is that this approach will likely yield a sub-optimal treatment regime if the regression model is misspecified.

Zhao et al. (2012) proposes a new method which formulates the problem as a misclassification problem and solves it by their Outcome Weighted Learning (OWL) approach. In our example, this method would predict whether the student attended tutoring or not for every student in the dataset based on their GPA and age. Each time the prediction is wrong, this error would be weighted by the outcome. Summing over all students yields then the total weighted classification error. The OTR is then defined as the treatment regime which minimizes that classification error. We refer readers, who find solving the OTR problem within a classification framework unintuitive, to Lemma 1 in the next section where we dwell on this further. Using non-parametric methods such as support vector machines to predict the treatment, their approach is robust compared to the one of Qian and Murphy (2011). However, the approach of Zhao et al. (2012) is formulated only for binary treatments, i.e., when there are only two treatments available. In our experience, we find that students often can choose from more than two treatment options, in particular our application identifying an OTR of student success programs for an introductory statistics course.

Another approach introduced by Tao and Wang (2017) called Adaptive Contrast Weighted Learning (ACWL) also perceives the estimation of the OTR as a misclassification problem. Tao and Wang (2017) estimates the OTR as the treatment regime which either minimizes the maximal (denoted as ACWL-$C_1$) or minimal (denoted as ACWL-$C_2$) expected loss if the subjects are assigned to a sub-optimal treatment, i.e., are misclassified. Applied to our example, we first predict a grade for every treatment for every student in the dataset. The method then assigns the treatment to the students which minimizes the respective loss for the whole student population if they were to be assigned a sub-optimal treatment. To make the grade prediction robust against misspecification of the model for the grade, Tao and Wang (2017) utilizes a so-called Augmented Inverse Probability-Weighted (AIPW) estimator for the grade. Their simulation study indicates that thereby the accuracy of the OTR estimation increases significantly.

We note that the two previous mentioned approaches both leverage methods which might resemble a "black box" to practitioners. In our example, the director of the tutoring center might be reluctant to exclude students from tutoring merely based on an OTR estimated by a misclassification error, a quantification that may be hard to grasp intuitively, or hard to justify to advisors and university administrators. To address this issue, Laber and Zhao (2015) introduces a novel tree-based reinforcement learning approach. The analysis leads to treatment regimes representable as a decision tree whose interpretability is widely appreciated. Those trees assign

treatments based on splits of the predictor space. In our example, this decision tree might recommend to assign tutoring hours just to students with a GPA of at most 3.5 as displayed in Fig. 1. To determine which split is optimal, i.e., which variable to use and which value shall be the split threshold, Laber and Zhao (2015) develops a new purity measure which assesses the optimality of the respective tree split. This measure, however, suffers from instability in certain circumstances, and we shall address these later. To this end, Tao et al. (2018) introduces a new, more robust purity measure based on the AIPW estimator developed in Tao and Wang (2017). Our contribution to the literature is threefold: first, we extend the latter two approaches to estimate OTR in an educational setting, including a study of their efficacy therein and considering of regression and random forest implementations not previously considered; second, we flush out details we found lacking in the literature and step through our R code (in the appendix) for practitioners to apply these approaches and make data-informed decisions relative to the OTR; third, we present, to our knowledge, the first quantitative analysis of potentially optimal combinations of Supplemental Instruction programs and Mathematics tutoring center visits for success in a core, bottleneck introductory STEM course.

The remaining parts of the paper are organized as follows. First, we define in mathematical terms what is meant precisely by an OTR and review the two methods developed by Laber and Zhao (2015) and Tao et al. (2018), respectively. Second, we show results of a simulation—where the OTR will be known—to scrutinize the performance of the two methods and their advantages and drawbacks. Finally, we apply the methods to determine an optimal "treatment cocktail" of Supplemental Instruction, Statistics tutoring, and Statistics recitation sections, all voluntarily chosen by

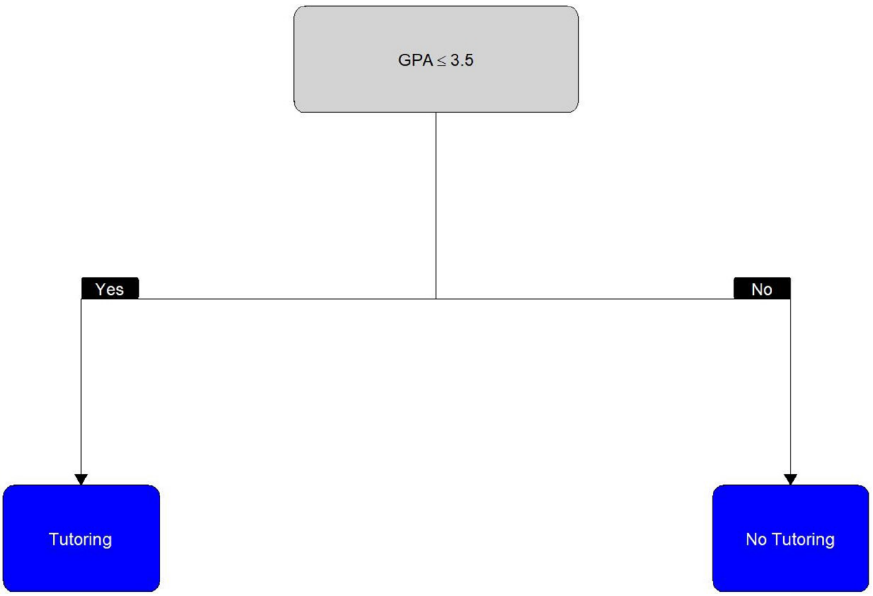


**Fig. 1** Example tree with a split at an GPA of 3.5.

students, for success in a large enrollment introductory statistics course. We also discuss the implications therein and more generally for estimating OTR of student success programs.

## 2 Methodology

To formalize the OTR problem, we consider a population with $n$ individuals. For each individual, we record $p$ characteristics in a vector $X = x_i \in \mathbb{R}^p$ where $i = 1, \ldots, n$. In addition, there are $m$ treatments $A_1, \ldots, A_m \in \mathcal{A}$ for each member of the population available where $\mathcal{A}$ denotes the set of all treatments. In this paper, we will consider only discrete treatments. A treatment regime $\pi$ is a function which assigns a treatment to each individual depending on its characteristics $X = x_i$, that is

$$\pi : \mathbb{R}^p \to \mathcal{A}, x_i \mapsto \pi(x_i).$$

Note that $\pi$ has no index $i$ as the treatment regime is for all members of the population identically. The potential outcome for an individual $i$ with characteristics $X = x_i$ under the treatment regime is then denoted as $Y^*\{\pi(x_i)\}$. Hence, we define an optimal treatment regime $\pi^{\text{opt}}$ as follows:

**Definition 1** We call a treatment regime $\pi^{\text{opt}}$ optimal if it satisfies

$$\pi^{\text{opt}} = \arg\max_{\pi \in \Pi} \mathbb{E}[Y^*\{\pi(X)\}], \tag{1}$$

where $\Pi$ denotes the class of all treatment regimes (Laber and Zhao 2015).

This definition reveals one major challenge in estimating the OTR: we neither observe the OTR nor the potential outcomes for all possible treatments as an individual can only be given one treatment. Hence, estimating the OTR directly proves difficult. Instead, we deduce a link between the potential and the observed outcomes to eventually use reinforcement learning-based decision trees to estimate the OTR.

Traditionally, decision trees are used for classification or regression problems where the quantity of interest is observed. For instance, you will observe the correct label $Y = y$ in a classification problem given certain characteristics $X = x$ and you can train your tree on these data. Yet, this does not hold when you estimate an OTR. To exemplify how we leverage the decision tree, nevertheless, we consider our example from the introduction with merely two treatments, i.e., whether the student shall attend tutoring or not. Identifying the treatment *no tutoring* with 0 and *tutoring* with 1, the set of possible treatments is $\mathcal{A} = \{0, 1\}$. Basically, a decision tree is a segmentation of the predictor space $\mathbb{R}^p$ of $X$. And so is the OTR $\pi^{\text{opt}}$. The OTR will divide the predictor space into two regions

$$\mathcal{R}_1^{\text{opt}} = \{x \in \mathbb{R} | \pi^{\text{opt}}(x) = 1\}$$

where students shall attend tutoring, i.e., $A = 1$ and in a region

$$\mathcal{R}_0^{\text{opt}} = \mathbb{R}^p \backslash \mathcal{R}_1 = \{x \in \mathbb{R} | \pi^{\text{opt}}(x) = 0\},$$

where students are not recommended to attend tutoring, i.e., $A = 0$. The regions are also referred to as rectangular regions or nodes. Hence, if we find a tree which—with the help of multiple splits based on certain characteristics—partitions the predictor space in the same way as the OTR does, we have found the OTR. The split of the predictor space associated with the tree in our example in Fig. 1 is displayed in Fig. 2 where every point represents a fictional student with the respective GPA and age. Based on a split at a GPA of 3.5, the tree partitions the predictor space into the rectangular region $\mathcal{R}_0$ where students shall not attend tutoring and $\mathcal{R}_1$ where students shall attend tutoring.

To estimate the OTR, we want to approximate the rectangular regions $\mathcal{R}_0^{\text{opt}}$ and $\mathcal{R}_1^{\text{opt}}$ for each split better or, as it is termed in the literature, to maximize our node purity. The notion of *purity* refers to a decision tree's goal of partioning the data into homogeneous groups. We will develop a purity measure in the next section.

### 2.1 Inverse probability-weighted purity measure

In this paper, we estimate the OTR for observational studies. Consequently, treatments in our sample data are not assigned but chosen by each individual. For every member of the sample, we record the individual's characteristic $X_i = x_i$, the treatment $A_i = a_i$ the individual chose, and the observed outcome $Y_i = y_i$ under $A_i = a_i$, where $i = 1, \ldots, n$ is our sample size. We assume that our observations are independent and drawn from the same distribution, i.e., $(X_i, A_i, Y_i)$ are *i.i.d.* To be able to identify for each individual $i$ with treatment $a$ its potential outcome $Y^*(a)$, we assume that the following holds:

**Assumption 1** (Stable unit treatment value) *For every individual* $i, j = 1, \ldots, n$ *and treatment* $a \in \mathcal{A}$, *it holds*

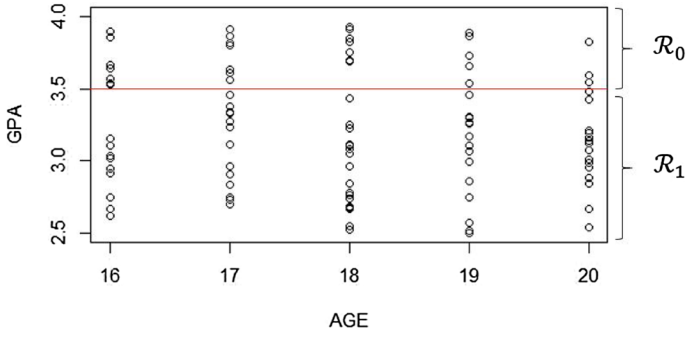


**Fig. 2** Rectangular region splits into two rectangles $\mathcal{R}_0$ and $\mathcal{R}_1$ based on a GPA of 3.5.

(i) *that there are no spill-over effects, i. e. the potential outcome of individual $i$ with treatment $a$ is not contingent on which treatment individual $j$, $j \neq i$, received; and*
(ii) *that there are no different versions of treatment $a$*

(Rubin 1980).

This is a standard assumption for causal inference. The two SUTVA implications might be violated in an educational environment for example if students work together or classes are taught by different teachers of varying quality. For the first implication, we thus must assume that there are no spill-over effects. The second implication should hold in our study as each treatment followed a strict template for instructional design and curricular content. The interested reader is referred to Sobel (2006) or VanderWeele and Hernan (2013) for causal inference if one of the implications does not hold.

It is important to understand that the observed outcome $Y$ and the potential outcome $Y^*$ are not necessarily the same. The potential outcome for an individual is the outcome which is realized under a specific treatment. Consequently, each treatment yields its own potential outcome. Yet, we assume with the following consistency assumption that we observe one of the potential outcomes:

**Assumption 2** (Consistency) *The observed outcome is the potential outcome that would be observed under the treatment received, that is*

$$Y = Y^*(A)$$

(Robins 1986).

To use our data for the estimation, we need two additional assumptions that are typically made for drawing causal inferences in observational studies. The first deals with the relationship between the potential outcome and the treatment:

**Assumption 3** (Strong ignorability) *All potential outcomes* $\{Y^*(a), a \in \mathcal{A}\}$ *are, conditioned on* $X$, *marginally independent from* $A$ (Gill and Robins 2001).

Assumption 3 means that in $X$, every necessary information for assigning treatments is observed, i.e., there are no confounding variables. Note that for observational studies, this assumption is unverifiable as mentioned by Zhang et al. (2012). In our student success studies, we thus must assume that we have collected enough student characteristics that there are no confounding variables remaining. We note that research on variables correlated to student success in higher education might help to collect relevant variables to assure that this assumption is reasonable. The interested reader is referred to Schneider and Preckel (2017) for a comprehensive literature review and a list of over a hundred variables like High School GPA (HSGPA) or socio-economic status which will be used in our study. We also note that in a study design, one should ensure to measure variables which are specific to potential student struggles in the scrutinized course, and not just connected to

academic achievement in general. For example, a student who performs well overall might nevertheless struggle with a statistics course.

The last necessary assumption requires that it is possible for every individual to choose any treatment option $a \in \mathcal{A}$:

**Assumption 4** (Positivity) *There exists an* $\epsilon > 0$, *such that* $P\{A = a|X\} \geq \epsilon$ *with probability 1 for all treatments* $a \in \mathcal{A}$ (Gill and Robins 2001).

Assumption 4 requires that every student has a chance to attend any student success program. This assumption is reasonable in our study as Supplemental Instruction, Statistics tutoring, and the Statistics recitation class are available to all students; students may attend any or all. Note that if treatments are assigned randomly, Assumptions 3 and 4 are automatically satisfied.

Equipped with these assumptions, we reformulate the definition of the OTR:

**Lemma 1** *If Assumptions 1 through 4 hold, we have*

$$\pi^{\text{opt}}(X) = \arg\max_{\pi \in \Pi} \mathbb{E}\left[\frac{Y}{P\{A = \pi(X)|X\}} \mathbb{1}_{\{A=\pi(X)\}}\right]. \tag{2}$$

Note that the proof is included in the appendix. Lemma 1 expresses the OTR in terms of the *observed* outcome $Y$ as prefigured in the beginning of this section. The probability $p_\pi(X) := P\{A = \pi(X)|X\}$ of choosing the treatment assigned by the treatment regime given certain characteristics is called the *propensity score*. Rosenbaum and Rubin (1983) shows that propensity scores help to account for self-selection bias in observational studies.

Equation (2) can be perceived as a weighted classification problem. Each time, a treatment regime $\pi$ assigns the observed treatment to an individual, i.e., classifies it correctly, it is weighted with the quotient of the observed outcome and the propensity score. The optimal treatment regime is then the one which maximizes this reward.

In a next step, we set out to minimize the variance of our OTR estimation. We prove in the appendix that substituting $Y - g(X)$ for $Y$ in (2) still leads to the OTR, for any function $g$. In addition, Laber and Zhao (2015) shows that by choosing $g(X) = \mathbb{E}[Y|X, A = \bar{\pi}]$ for any treatment regime $\bar{\pi}$, the thereby obtained expectation can be estimated with minimal variance. As we are aiming at the OTR, it is natural that we choose the treatment regime $\bar{\pi}$ to be the optimal treatment regime $\pi^{\text{opt}}$. This leads us to the estimator

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\{y_i - \hat{m}(x_i)\} \mathbb{1}_{\{a_i=\pi(x_i)\}}}{p_\pi(x_i)}, \tag{3}$$

where $m(x_i) := \mathbb{E}[Y|X = x_i, A = \pi^{\text{opt}}(x_i)]$ denotes the conditional mean outcome of student $i$ under the OTR given their characteristics $x$, whereas $y_i$ is the student's observed outcome. To obtain $\hat{m}(x)$, we estimate for all treatment options $a \in \mathcal{A}$ and

all characteristics $X = x$ the conditional mean outcome $\mu_a(x) = \mathbb{E}[Y|X = x, A = a]$. We then set $\hat{m}(x) = \max_{a \in \mathcal{A}} \hat{\mu}_a(x)$.

The estimator in Eq. (3) lays the foundation for the *purity measure* of our tree growing algorithm. Denote the treatment rule which assigns treatment $a$ to all individuals with characteristics $x$ in some arbitrary rectangle $r$ and treatment $a'$ otherwise with $\pi_{r,a,a'}$. Then, the purity of the split of the rectangular region $\mathcal{R}$ according to that treatment rule is measured as follows:

$$\begin{aligned}\mathcal{P}^{LZ}(\mathcal{R}, r) = &\max_{a,a' \in \mathcal{A}} \left\{ \sum_{i=1}^{n} \frac{\{y_i - \hat{m}(x_i)\} \mathbb{1}_{\{x_i \in \mathcal{R}\}} \mathbb{1}_{\{a_i = \pi_{r,a,a'}(x_i)\}}}{p_{\pi_{r,a,a'}}(x_i)} \right\} \\ &\cdot \left\{ \sum_{i=1}^{n} \frac{\mathbb{1}_{\{x_i \in \mathcal{R}\}} \mathbb{1}_{\{a_i = \pi_{r,a,a'}(x_i)\}}}{p_{\pi_{r,a,a'}}(x_i)} \right\}^{-1},\end{aligned} \tag{4}$$

where $n$ denotes our sample size. The numerator is basically inherited from Eq. (3). Note that the indicator function $\mathbb{1}_{\{x_i \in \mathcal{R}\}}$ ensures that the purity is estimated only based on students in the rectangular region $\mathcal{R}$ which we would like to split. The denominator is introduced to stabilize the purity measure in case that the assigned treatment is not often observed for the subjects in the node to split. As multiple treatment assignments $a$ and $a'$ are conceivable for each split, the combination which maximizes the estimator in (4) is identified as the purity associated to that split.

Note that the purity measure does not necessarily lead to an OTR according to Definition 1 as it finds only the OTR for each individual node; this point is criticized by Doubleday et al. (2018). Another drawback of the purity measure defined in (4) is specific to observational studies. As individuals assign themselves the treatment they receive, we do not know the probability that they will choose a certain treatment $a$ based on their characteristics. Hence, we need to estimate the propensity scores $p_a$. As Tao et al. (2018) points out, this makes the purity measure vulnerable to misspecification of the propensity score model. Therefore, they introduced a so-called augmented inverse probability-weighted (AIPW) purity measure which we will detail in the next section.

### 2.2 Augmented inverse probability-weighted purity measure

Tao et al. (2018) proposes to use the estimator developed by Bang and Robins (2005)

$$\hat{\mathbb{E}}\{Y^*(a)\} = \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}_{i,a}^{AIPW}(x_i) = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i \cdot \mathbb{1}_{\{a_i = a\}}}{\hat{p}_a(x_i)} + \left(1 - \frac{\mathbb{1}_{\{a_i = a\}}}{\hat{p}_a(x_i)}\right) \hat{\mu}_a(x_i) \tag{5}$$

for the mean of the potential outcomes $\mathbb{E}\{Y^*(a)\}$ where $\hat{p}_a(x_i)$ denotes the estimated propensity score and $\hat{\mu}_a(x_i)$ denotes the estimator of the conditional mean outcome given characteristics $x_i$ and treatment $a$. The first summand in Eq. (5) is drawn from the expectation in Lemma 1. The second one can be thought of as a stabilizer for students in our dataset where the observed treatment $a_i$ is not the treatment $a$ we

would like to estimate the potential mean outcome for. In this case, the first summand in (5) vanishes as $\mathbb{1}_{\{a_i=a\}} = 0$. Instead, the estimated conditional mean outcome $\hat{\mu}_a$ is taken into account for this student.

Of importance, Tao et al. (2018) shows that if either the propensity score model $p_a(X)$ or the conditional mean model $\mu_a(X)$ is correctly specified, $\hat{\mathbb{E}}\{Y^*(a)\}$ in (5) is a consistent estimator of $\mathbb{E}\{Y^*(a)\}$, a desired result as the sample size increases.

Based on this estimator, Tao et al. (2018) proposes the following *purity measure:*

$$\mathcal{P}^{\text{Tao}}(\mathcal{R}, r) = \max_{a_1,a_2\in\mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \sum_{a\in\mathcal{A}} \hat{\mu}_{i,a}^{AIPW}(x_i) \mathbb{1}_{\{\pi_{r,a_1,a_2}(x_i)=a\}} \mathbb{1}_{\{x_i\in\mathcal{R}\}}. \tag{6}$$

Like the purity measure $\mathcal{P}^{LZ}$, $\mathcal{P}^{\text{Tao}}$ identifies the combination of treatments $a_1$ and $a_2$ which maximizes the estimator in (6) as the purity associated to the respective split. Summing over all treatment options $a \in \mathcal{A}$ ensures that we merely consider the AIPW estimator of the conditional mean outcome given the assigned treatment under the candidate treatment regime. Consequently, the purity measure $\mathcal{P}^{\text{Tao}}$ averages the respective AIPW estimator for the respective treatment for every student in the rectangular region $\mathcal{R}$, which is a consistent estimator of $\mathbb{E}[Y^*\{\pi(X)\}]$. Equipped with the two purity measures, we will elaborate on how to grow a tree in the next section.

### 2.3 Growing a tree

Consider a rectangular region $\mathcal{R}$ like displayed in Fig. 2 which we would like to split. Of course, we set out to find the split which is the most pure in terms of our purity measures. Yet, as our purity measures become more instable for smaller nodes—especially in the case of $\mathcal{P}^{LZ}$ in (4) where we omit observations when they do not exhibit a compatible treatment—we impose a minimum node size $\gamma$ to even consider a split with the following two rules:

**Rule 1** *If the number of observations in our sample which fall into the rectangular region $\mathcal{R}$ only allows for child nodes which undercut a minimum node size of $\gamma$, i.e. $\sum_{i=1}^{n} \mathbb{1}_{\{x_i\in\mathcal{R}\}} < 2\gamma$, do not split* (Tao et al. 2018).

**Rule 2** *If every possible split yields a child node with size smaller than the minimum node size $\gamma$, we do not split the node* (Tao et al. 2018).

However, this is not enough to split the rectangular region $\mathcal{R}$ for sure. With each split, not only our purity increases but our tree complexity as well. An increased complexity potentially deteriorates our tree's interpretability and lets the tree resemble a black-box to practitioners. We introduce two rules to address this issue. First, we set a maximum depth, i.e., the maximum number of layers for a tree:

**Rule 3** *If a pre-defined maximum tree depth is attained, the tree growing is stopped* (Tao et al. 2018).

Furthermore, we define a minimum purity gain at which we are willing to tolerate the disadvantages of a more complex tree:

**Rule 4** *If the number of observations in the rectangular region exceeds* $2\gamma$ *and the maximum depth of the tree is not attained yet, compute the rectangle r which is associated with the maximal purity split. If the purity gain exceeds a pre-defined threshold* $\lambda > 0$, *i.e.,* $\mathcal{P}(\mathcal{R}, r) \geq \mathcal{P}(\mathcal{R}) \cdot (1 + \lambda)$, *then split the rectangular region* $\mathcal{R}$ *into the rectangle r and its complement* $r^c$, *otherwise do not split* (Tao et al. 2018).

There is no consensus in the literature on how to choose the minimum purity gain $\lambda$. Laber and Zhao (2015) proposes to define $\lambda$ by expert judgement to grow a tree and later prune it back using another purity threshold estimated by a cross-validation approach. However, Tao et al. (2018) suggests to directly estimate $\lambda$ via maximizing the tenfold cross-validation estimator of the mean of the potential outcomes, which we will elaborate on in the next section. But first, we will summarize the outlined steps in an algorithm.

To compute the purity measures $\mathcal{P}^{LZ}$ in (4) or $\mathcal{P}^{\text{Tao}}$ in (6), we need to find estimates for the propensity scores $\hat{p}_a(x_i)$ and for the conditional mean outcomes $\hat{\mu}_a(x_i)$s. Our algorithm estimates the propensity scores via multinomial regression and the conditional mean outcomes either using random forest or linear regression. All estimates are based on the full data. This yields the following algorithm:

**Step 1** Set values for the minimum purity gain $\lambda$, the minimum node size $\gamma$, and the maximum tree depth.

**Step 2** Obtain estimates for propensity scores $\hat{p}$ and conditional mean outcomes $\hat{\mu}_a$.

**Step 3** Compute either $\hat{m}(x_i)$ for any individual $i$ for the Laber and Zhao (2015) approach or $\hat{\mu}^{AIPW}_{i,a}(x_i)$ for $a \in \mathcal{A}$ and $i = 1, \ldots, n$ for the Tao et al. (2018) approach and set $l = 1$.

**Step 4** At every rectangle $\mathcal{R}_l$, check the four *Rules* 1 to 4. If any of them is satisfied, do not split. If $l = 1$, find the best single treatment for all subjects in $\mathcal{R}_l$ by

$$\arg\max_{a \in \mathcal{A}} \left\{ \sum_{i=1}^{n} \frac{\{y_i - \hat{m}(x_i)\} \mathbb{1}_{\{x_i \in \mathcal{R}_l\}} \mathbb{1}_{\{a_i = a\}}}{\hat{p}_a(x_i)} \right\} \cdot \left\{ \sum_{i=1}^{n} \frac{\mathbb{1}_{\{x_i \in \mathcal{R}_l\}} \mathbb{1}_{\{a_i = a\}}}{\hat{p}_a(x_i)} \right\}^{-1}$$

for the Laber and Zhao (2015) approach or

$$\arg\max_{a \in \mathcal{A}} \frac{1}{n} \sum_{i=1}^{n} \hat{\mu}^{AIPW}_{i,a}(x_i) \mathbb{1}_{\{x_i \in \mathcal{R}_l\}}$$

for the Tao et al. (2018) approach.

Otherwise, compute the rectangle $r$ which is associated with the maximal purity split using purity measure $\mathcal{P}^{LZ}$ in (4) or $\mathcal{P}^{\text{Tao}}$ in (6) and split the rectangular region $\mathcal{R}_l$ into $r$ and $r^c$.

**Step 5** Set $l = l + 1$ and repeat Step 4 until all nodes are terminal.

## 3 Simulation study

In reality, we do not know what treatment is optimal for a subject. We only observe the treatment choice of the individual and the associated outcome. This makes assessing the performance of the introduced methods for real-world problems impossible. However, in this section, we conduct a simulation study in which we do know the underlying structure of the outcome. This equips us with knowledge about the optimal treatment regime (OTR). In this case, an intuitive performance measure of our methods is the percentage of correctly assigned OTRs which we denote by `opt%`. To test the stability of the methods, we misspecify the models for the propensity score $\hat{p}$ and the conditional mean outcome $\mu_a$. This section illustrates the implementation of this approach for which we utilized version 3.6.3 of `R` as the programming software R Core Team (2020). The interested reader can find further details and an example in Appendix B.

Our approach for the simulation study is partly based on Tao et al. (2018). We simulate two data sets. The first dataset contains five covariates $X_1, \ldots, X_5$ which are all normally distributed with mean zero and variance one. The second dataset consists of 21 covariates where the first ten covariates are independently normally distributed with mean zero and variance one. The normal distributions in both models represent continuous variables one sees in practice. We did not directly model them after our particular application to ensure that our simulation results do apply in a more general practice. Ten covariates are Bernoulli distributed, i.e., $X_{11}, \ldots X_{14} \sim \text{Ber}(0.25)$; $X_{15}$, $X_{16} \sim \text{Ber}(0.4)$; $X_{17} \sim \text{Ber}(0.1)$; $X_{18} \sim \text{Ber}(0.3)$; $X_{19} \sim \text{Ber}(0.6)$; and $X_{20} \sim \text{Ber}(0.8)$. The success probabilities were chosen to be akin to those of one of the empirical distributions of the binary variables in our application. The remaining covariate $X_{21}$ is an unordered factor variable with six levels where the levels are attained with the probabilities

$$P(X_{21} = i) = \begin{cases} 0.31 & \text{if } i = 1 \\ 0.41 & \text{if } i = 2 \\ 0.16 & \text{if } i = 3 \\ 0.04 & \text{if } i = 4 \\ 0.07 & \text{if } i = 5 \\ 0.01 & \text{if } i = 6. \end{cases}$$

The probabilities are chosen, so that they resemble the empirical distribution of the GPA variables in our application later. In each dataset, there are three treatments available, i.e., the set of all treatments is $\mathcal{A} = \{0, 1, 2\}$. The probability that a certain treatment is observed given the characteristics $X$ is determined by

$$P\{A = 0|X\} = \left(1 + \mathrm{e}^{0.5X_1 + 0.5X_4} + \mathrm{e}^{-0.5X_1 + 0.5X_5}\right)^{-1}, \tag{7}$$

$$P\{A = 1|X\} = \mathrm{e}^{0.5X_1 + 0.5X_4} \cdot \left(1 + \mathrm{e}^{0.5X_1 + 0.5X_4} + \mathrm{e}^{-0.5X_1 + 0.5X_5}\right)^{-1}, \tag{8}$$

$$P\{A = 2|X\} = 1 - P\{A = 0|X\} - P\{A = 1|X\}. \tag{9}$$

We define our underlying OTR to be

$$\pi^{\text{opt}}(X) = \begin{cases} 0 & \text{if } X_1 \leq 0, X_2 \leq 0.5 \\ 2 & \text{if } X_1 > 0, X_3 \leq 0.5 \\ 1 & \text{otherwise.} \end{cases} \tag{10}$$

For our first dataset, we generate the associated observed outcome given the characteristics $X$, the simulated treatment $A$, and the defined OTR $\pi^{\text{opt}}$ by

$$\begin{aligned} Y = 0.79 + X_4 + X_5 + 2\mathbb{1}_{\{A=0\}}(2\mathbb{1}_{\{\pi^{\text{opt}}(X)=0\}} - 1) \\ + 1.5\mathbb{1}_{\{A=2\}}(2\mathbb{1}_{\{\pi^{\text{opt}}(X)=2\}} - 1) + \epsilon. \end{aligned} \tag{11}$$

For our second dataset, we utilize a more sophisticated model for generating the outcome

$$\begin{aligned} Y = 0.79 + X_4 + X_5 + \sin(X_8) + X_{12} + X_6 \cdot X_7 + 2\mathbb{1}_{\{A=0\}}(2\mathbb{1}_{\{\pi^{\text{opt}}(X)=0\}} - 1) \\ + 1.5\mathbb{1}_{\{A=2\}}(2\mathbb{1}_{\{\pi^{\text{opt}}(X)=2\}} - 1) + \epsilon. \end{aligned} \tag{12}$$

The random variable $\epsilon$ in Eqs. (11) and (12) follows a standard normal distribution.

For both models, we are also interested in the potential mean outcome $\mathbb{E}[Y^*\{\pi^{\text{opt}}(X)\}]$. In the first model, it can be shown by an easy computation that the mean equals two in this case. For the second model, it is possible to deduce a mean of 2.25 theoretically by leveraging the symmetry of the sine function with respect to the origin and the independence of the normal variates.

To test our approach's robustness against misspecification of the model for propensity scores $\hat{p}$, we consider two models. The first one includes every covariate which influences the probability of treatment choice and is therefore correctly specified as

$$\log(p_d/p_0) = \beta_{0d} + \beta_{1d}X_1 + \beta_{2d}X_4 + \beta_{3d}X_5, \quad d = 1, 2.$$

The second one includes only $X_2$ and $X_3$ which do not determine the probability of treatment choice. The model is therefore misspecified as

$$\log(p_d/p_0) = \beta_{0d} + \beta_{1d}X_2 + \beta_{2d}X_3, \quad d = 1, 2.$$

To check the sensitivity regarding the choice of model for the conditional mean outcome $\mu_a = \mathbb{E}[Y|X, A = a]$, we will consider both linear regression-based and random forest-based estimation procedures which we will misspecify by omitting the fourth covariate $X_4$. By considering linear regression (REG) and random forest (RF), we check the sensitivity of our approaches regarding the statistical procedure to estimate $\mu_a$, and an aspect Tao et al. (2018) and Laber and Zhao (2015) did not consider. They used either linear regression or random forest respectively.

We set the minimal node size $\gamma$ to 5% of all data points in our training set and the maximal depth of the tree equals 5 as Tao et al. (2018) proposes in their simulation approach.

Finally, to perform the algorithm in Sect. 2, we need to choose the required minimum purity gain $\lambda$ for a split to be performed. Following the approach of Tao et al. (2018), we divide a dataset of 1000 simulations of the model in (11) ten times randomly in ten subsets for every $\lambda$ we seek to evaluate. For each repetition $j$, we use nine of them to estimate the tree using the purity measure $\mathcal{P}^{\text{Tao}}$ in Eq. (6). For the remaining subset, we predict with the estimated tree the respective OTR for each observation. We then estimate the mean potential outcome $\mathbb{E}[Y^*(\hat{\pi}^{\text{opt}})]$ via Eq. (5). Finally, we average the thereby acquired values to get an average mean potential outcome associated with each $\lambda$ value. Consequently, we choose the $\lambda$ which yields the highest mean on average. This approach yields an optimal $\lambda$ of 0.02—for which we could reproduce the results in the first simulation study of Tao et al. (2018)—as a required minimal purity gain at each step. We set this $\lambda$ for *both* models to make sure that performance differences can only be attributed to the different models and not to a different $\lambda$.

The results of the next section are based on training datasets of size 1000 and on test datasets of size 500. We repeated the simulation 500 times. For every simulation run, we simulated the covariates according to the specified distributions. Based on them, we assigned treatments in the training set according to Eqs. (7) to (9). We then determined the true OTR $\pi^{\text{opt}}$ for every individual based on (10). The observed outcome was computed by either Eq. (11) or (12). On the thereby completed training set, we trained the propensity score model and the conditional mean outcome model to eventually grow the tree either maximizing $\mathcal{P}^{\text{Tao}}$ or $\mathcal{P}^{LZ}$. Subsequently, we leveraged this tree to predict the OTR for each individual in the test data set and assigned this as the individual's treatment. Comparing the predicted OTR with the true one, we compute `opt%`. Furthermore, we deduced the observed outcome under the predicted OTR for every individual in the test set by leveraging either Eq. (11) or (12). Due to consistency of observed and potential outcome postulated in Assumption 2, we can simply average the observed outcomes under the predicted OTR to obtain the mean potential outcome in the test data set. The thereby acquired results for the approach by Tao et al. (2018) and Laber and Zhao (2015) are presented in Tables 1 and 2.

For the linear model in Eq. (11), the results are displayed in Table 1. When using the purity measure $\mathcal{P}^{\text{Tao}}$ proposed by Tao, we observe that the method works best if linear regression is used to estimate the conditional mean outcomes $\mu_a$. Furthermore, the method seems to be rather immune against misspecification of either the propensity score model $\hat{p}$ or the model for $\mu_a$, as expected from the development in Sect. 2.2. However, if both models are misspecified, we record a significant decline in the mean of correctly assigned treatments `opt%` along with a higher standard deviation. If we use the model of Tao et al. (2018) along with a random forest approach to grow the tree, we observe worse results especially if we misspecify the propensity score model, as well. This is in line with the literature as Bang and Robins (2005) deduces the AIPW estimator in Eq. (5) assuming explicitly a parametric outcome regression model. Using a random forest, the AIPW estimator apparently

**Table 1** Simulation results with three treatment options, five covariates, and outcome generated according to Eq. (11).

| $\mathbb{E}[\boldsymbol{Y}^*\{\boldsymbol{\pi}^{\text{opt}}(\boldsymbol{X})\}] = 2$ | | | | |
|---|---|---|---|---|
| $\hat{p}$ | $\mu_a$ | Method | Model in Eq. (11) | |
| | | | opt% | $\hat{\mathbb{E}}[\boldsymbol{Y}^*(\hat{\boldsymbol{\pi}}^{\boldsymbol{opt}})]$ |
| Correct | REG correct | Tao | 0.98 (0.02) | 1.97 (0.09) |
| | REG incorrect | Tao | 0.97 (0.03) | 1.95 (0.10) |
| Incorrect | REG correct | Tao | 0.96 (0.05) | 1.95 (0.12) |
| | REG incorrect | Tao | 0.84 (0.17) | 1.74 (0.29) |
| Correct | RF correct | Tao | 0.88 (0.12) | 1.81 (0.22) |
| | RF incorrect | Tao | 0.91 (0.07) | 1.85 (0.14) |
| Incorrect | RF correct | Tao | 0.67 (0.08) | 1.48 (0.17) |
| | RF incorrect | Tao | 0.60 (0.10) | 1.37 (0.21) |
| Correct | REG correct | Laber | 0.97 (0.04) | 1.94 (0.10) |
| | REG incorrect | Laber | 0.96 (0.03) | 1.94 (0.10) |
| Incorrect | REG correct | Laber | 0.92 (0.10) | 1.87 (0.19) |
| | REG incorrect | Laber | 0.91 (0.11) | 1.84 (0.20) |
| Correct | RF correct | Laber | 0.96 (0.05) | 1.93 (0.12) |
| | RF incorrect | Laber | 0.96 (0.04) | 1.93 (0.10) |
| Incorrect | RF correct | Laber | 0.90 (0.11) | 1.83 (0.20) |
| | RF incorrect | Laber | 0.88 (0.13) | 1.81 (0.24) |

The simulation is run with 500 replications, i.e., $n = 500$. The size of the training set is 1000, whereas the size of the test set is 500. The minimal node size required is 50 and the maximal depth 5. The minimal required purity gain $\lambda$ for every split was set to 0.02. The letter $\hat{p}$ denotes the propensity score model and $\mu_a$ displays which model was used to estimate the conditional mean outcomes. The abbreviations REG and RF denote whether a linear regression or a random forest was used for $\mu_a$. The abbreviation opt% denotes the empirical mean of the percentage of subjects correctly assigned to their true optimal treatments. The respective standard deviation is given in parentheses. Finally, note that the true value of $\mathbb{E}[Y^*\{\pi^{\text{opt}}(X)\}]$ is 2

loses its double robustness and thereby becomes vulnerable regarding the misspecification of the propensity score model. Chernozhukov et al. (2018) delivers a possible explanation. They argue that non-parametric estimators for nuisance parameters such as the conditional mean outcome in (5) might fail to satisfy the so-called *Donsker properties*. In such a case, plugging the non-parametric estimator naively into estimation equations like (5) induces significant bias of the thereby estimated quantity of interest. They propose *Neyman-orthogonal moments* and *crossfitting of the nuisance estimators* to remedy this issue. We refer the interested reader to their paper for further details as this is beyond the scope of this paper.

The results obtained by the method of Laber and Zhao (2015) generally coincide with the method of Tao et al. (2018). The method of Laber and Zhao (2015) also works better for linear regression as the model for $\mu_a$. However, while the method of Tao et al. (2018) yields slightly better results when using linear regression

**Table 2** Simulation results with three treatment options, 21 covariates, and outcome generated according to Eq. (12)

| $\mathbb{E}[\boldsymbol{Y}^*\{\boldsymbol{\pi}^{\text{opt}}(\boldsymbol{X})\}] = 2.25$ | | | | |
|---|---|---|---|---|
| $\hat{p}$ | $\mu_a$ | Method | Model in Eq. (12) | |
| | | | opt% | $\hat{\mathbb{E}}[\boldsymbol{Y}^*(\hat{\boldsymbol{\pi}}^{\text{opt}})]$ |
| Correct | REG correct | Tao | 0.95 (0.05) | 2.18 (0.13) |
| | REG incorrect | Tao | 0.92 (0.08) | 2.12 (0.16) |
| Incorrect | REG correct | Tao | 0.94 (0.06) | 2.16 (0.13) |
| | REG incorrect | Tao | 0.76 (0.18) | 1.88 (0.31) |
| Correct | RF correct | Tao | 0.79 (0.17) | 1.91 (0.31) |
| | RF incorrect | Tao | 0.65 (0.20) | 1.65 (0.38) |
| Incorrect | RF correct | Tao | 0.67 (0.11) | 1.72 (0.22) |
| | RF incorrect | Tao | 0.47 (0.12) | 1.37 (0.26) |
| Correct | REG correct | Laber | 0.94 (0.07) | 2.14 (0.16) |
| | REG incorrect | Laber | 0.93 (0.08) | 2.12 (0.17) |
| Incorrect | REG correct | Laber | 0.87 (0.12) | 2.06 (0.23) |
| | REG incorrect | Laber | 0.85 (0.13) | 2.00 (0.24) |
| Correct | RF correct | Laber | 0.90 (0.11) | 2.07 (0.23) |
| | RF incorrect | Laber | 0.85 (0.15) | 1.98 (0.28) |
| Incorrect | RF correct | Laber | 0.84 (0.14) | 1.98 (0.26) |
| | RF incorrect | Laber | 0.56 (0.22) | 1.50 (0.41) |

The simulation is run with 500 replications, i.e., $n = 500$. The size of the training set is 1000, whereas the size of the test set is 500. The minimal node size required is 50 and the maximal depth 5. The minimal required purity gain $\lambda$ for every split was set to 0.02. The letter $\hat{p}$ denotes the propensity score model and $\mu_a$ displays which model was used to estimate the conditional mean outcomes. The abbreviations REG and RF denote whether a linear regression or a random forest was used for $\mu_a$. The abbreviation opt% denotes the empirical mean of the percentage of subjects correctly assigned to their optimal treatments. The respective standard deviation is given in parentheses. Finally, note that the true value of $\mathbb{E}[Y^*\{\pi^{\text{opt}}(X)\}]$ is 2.25

to estimate the conditional mean outcome, the method of Laber and Zhao (2015) does not exhibit the significant decline when changing to a random forest. This is also in line with the literature as Laber and Zhao (2015) deduces its purity measure independently of the choice between parametric and non-parametric outcome regressions.

Moving from our rather simple model for the observed outcome in Eq. (11) to the more sophisticated model of Eq. (12), we observe a similar pattern in the results of Table 2. In general, the estimates for opt% worsen with lower mean and higher standard deviation which we would anticipate for a more complicated model. For both methods, the regression model still works best for estimating the conditional mean outcome $\mu_a$. Note that the method Laber and Zhao (2015) now presents a steep performance decline when the model for both propensity score and conditional mean outcome is misspecified and a random forest is used. According to our

simulation results, we shall use a linear regression model for $\mu_a$ for both methods in the application in the next section.

## 4 Results: optimal treatment regime for student success in introductory statistics

One of the courses at San Diego State University (SDSU) with the highest failure rate is an introductory statistics course for undergraduates. To improve student success in this course, the university has implemented three programs in recent years

- *Statistics recitation class* A weekly one-unit supplemental course taught by graduate teaching assistants. The sessions review that week's lecture content and provide in-class assignments to encourage students to practice the statistics material. Upon successfully passing this one-unit class, students receive an additional two percentage points on their overall grade in the introductory statistics course. To pass, students need to attend a certain number of in-class meetings and complete all assignments.
- *Supplemental instruction (SI)* Weekly active problem solving sessions taught by a student who earned an A- or better in a recent semester of the introductory statistics course. The sessions are in a classroom setting with one SI leader. Class size depends on how many students voluntarily choose to attend. The SI program follows the peer-assisted tutoring model of Martin and Arendale (1992), and SI leaders receive extensive training on how to lead SI sessions within this model.
- *Tutoring at SDSU's Math & Stats Learning Center (MSLC)* One-on-one tutoring by qualified students or graduate teaching assistants available for certain hours during the week.

All three programs are voluntary: students sign up for the one-unit statistics recitation class and attend as many SI and MSLC tutoring sessions as they wish each week of the semester. Each individual program is potentially costly, both in terms of space and personnel, depending on attendance. The administration is thus interested in an optimal success program for each student among these three options not only to maximize student performance in the course, but appropriately allocate resources. Therefore, the goal of this section is to deduce an individual recommendation based on the student's characteristics—or to put in the terms of Sect. 2—to deduce an optimal treatment regime. The data which we use for this analysis are summarized in the next subsection.

### 4.1 Description of data

The data set covers enrollment in introductory statistics for two spring semesters with the same instructor; 1397 students in total. We did not include graduate nor transfer students. If students repeated the course multiple times due to a failing grade in previous semesters (96 students), we took only the most recent enrollment into

account. Our response variable $Y$ is *Grade*, which is the number of points scored out of 1030 possible points. This course grade is adjusted for the two percentage point boost if the student passed the supplemental statistics recitation class.

Table 3 displays the available *treatments*, i.e., the combination of the programs a student attended, and how often the respective combination was observed. A majority of 58% of students participated in at least one program and every fifth student attended at least two programs. For a student to be counted as "participating" or "attending" the SI or MSLC, it is sufficient to have participated in one SI or tutoring session, respectively. This seems to be a low threshold; however, a higher one would necessarily diminish the already low percentages of treatments 5, 6, 7, and 8 further making the statistical analysis potentially infeasible. Besides the response variable *Grade* and the treatment variable, there are 19 covariates which consider only information available at the beginning of each semester. Table 4 gives an overview of the categorical variables including the number of levels and the proportion of students for each of them. Certain variables like *Gender* are easy to understand, whereas others require further explanation. We consider, for instance, a student to be the *First Generation at College* if the student's father and mother have at most *some* college education, which means that they attended some college courses without graduating with a degree. The *COMPACT Scholarship* is awarded to students from a local school district to foster student success from this area. The three GPAs have been adjusted for coursework during the observed semester to make sure that they include only information available at the beginning of the semester. The level *N/A* is necessary to distinguish between students who have a GPA of 0.00 due to weak performance and students who simply did not do any course work yet. For the *Total GPA* and the *Campus GPA*, this distinction seems to be rather irrelevant as less than 1% of the students fall in one of these categories. However, we consider two spring semesters in our study. If we were to include fall semesters, this distinction would gain importance as in the fall semester the number of first-time-freshmen is considerable.

Table 5 gives an overview of the included continuous variables. Note that ACT scores were converted to SAT scores by the official converting formulas. Converted

**Table 3** Summary of the prevalence of treatments in the given dataset

| Level | Description | Proportion of students (%) |
|---|---|---|
| 1 | No attendance: Stat recitation, MSLC, SI | 42 |
| 2 | No attendance: Stat recitation, SI; attendance: MSLC | 9 |
| 3 | No attendance: Stat recitation MSLC; attendance: SI | 11 |
| 4 | No attendance: SI, MSLC; attendance: Stat recitation | 16 |
| 5 | No attendance: SI; attendance: Stat recitation, MSLC | 8 |
| 6 | No attendance: MSLC; attendance: SI, Stat recitation | 4 |
| 7 | No attendance: Stat recitation; attendance: SI, MSLC | 5 |
| 8 | Attendance: Stat recitation, MSLC, SI | 5 |

**Table 4** Summary of the categorical variables in the given dataset

| Variable | Description | Proportion of students (%) |
|---|---|---|
| Gender | 0 is female | 60 |
| | 1 is male | 40 |
| URM | 0 is no member of URM | 60 |
| underrepresented minorities | 1 is member of URM | 40 |
| AP credit | 0 is no AP credit | 42 |
| | 1 is AP credit | 58 |
| First generation at college | 0 is not first generation college | 77 |
| | 1 is First Generation College | 23 |
| Compact scholar | 0 is Compact scholar | 89 |
| | 1 is not compact scholar | 11 |
| Science, technology, | 0 is no STEM major | 69 |
| engineering, mathematics (STEM) | 1 is STEM major | 31 |
| Premajor | 0 is major declared | 18 |
| | 1 is no major declared yet | 82 |
| Total GPA | A is GPA in [3.3, 4] | 32.0 |
| | B is GPA in [2.3, 3.3) | 49.7 |
| | C is GPA in [1.3, 2.3) | 14.5 |
| | D is GPA in (0.00, 1.3) | 3.3 |
| | F is GPA of 0.00 | 0.1 |
| | N/A is no course work yet | 0.4 |
| Campus GPA | A is GPA in [3.3, 4] | 31.3 |
| | B is GPA in [2.3, 3.3) | 48.3 |
| | C is GPA in [1.3, 2.3) | 15.9 |
| | D is GPA in (0.00, 1.3) | 3.7 |
| | F is GPA of 0.00 | 0.1 |
| | N/A is no campus course work yet | 0.8 |
| Transfer GPA | A is GPA in [3.3, 4] | 14.3 |
| | B is GPA in [2.3, 3.3) | 7.7 |
| | C is GPA in [1.3, 2.3) | 1.1 |
| | D is GPA in (0.00, 1.3) | 0.1 |
| | F is GPA of 0.00 | 0.1 |
| | N/A is no transferred course work | 76.7 |

SAT Math and SAT Verbal scores are not bound to add up to the respective converted SAT composite score as the underlying ACT composite score is the average and *not* the sum of the ACT scores of both sections. Therefore, the average SAT Composite score in Table 5 is not the sum of the SAT Math and Verbal score. If neither ACT-Scores nor SAT-Scores were provided, the student was not included in the dataset (81 students were omitted). The variables *Campus units earned* and *Total units earned* are the respective number of units at the beginning of the observed

**Table 5** Summary of the continuous variables in the given dataset

| Variable | Min | Max | Mean | Median | SD |
|---|---|---|---|---|---|
| Grade | 24 | 1018 | 776 | 812 | 174 |
| Age | 17 | 25 | 19 | 18 | 1 |
| SAT composite | 730 | 1540 | 1169 | 1180 | 131 |
| SAT math | 320 | 790 | 582 | 580 | 74 |
| SAT verbal | 280 | 780 | 498 | 490 | 110 |
| High school GPA | 2.40 | 4.44 | 3.69 | 3.73 | 0.31 |
| Transfer units accepted | 0 | 49 | 2 | 0 | 5 |
| Campus units earned | 0 | 130 | 21 | 16 | 15 |
| Total units enrolled in semester | 3 | 23 | 15 | 15 | 2 |
| Total units earned | 0 | 137 | 30 | 24 | 18 |

semester. Furthermore, the variable *Total units enrolled in semester* does not include the one unit which students receive upon successfully passing the statistics recitation class. The next subsection deals with the application of the two proposed methods to the introduced data.

### 4.2 Estimation of the optimal treatment regime

For both trees presented in this section, a minimum purity gain of $\lambda = 0.025$ at each split was required. The conditional mean outcome $\mu_a$ was—taking our simulation results into account—estimated in both cases via a linear regression where all covariates, the recorded treatment, and their interaction terms were included. The potential average grade $\mathbb{E}[Y^*\{\pi^{\text{opt}}(X)\}]$ was estimated as presented in Eq. (5).

Applying the purity measure $\mathcal{P}^{\text{Tao}}$ defined in (6) gives us the tree displayed in Fig. 3. It recommends for students with at most 14 units earned on campus to attend all programs, i.e., enroll in the statistics recitation class and attend MSLC and SI at least once (node 1). This might be reasonable as students which are new to the university system often do not know which format works best for them. Hence, enrolling in the statistics recitation class and attending the other two programs at least once could facilitate their decision-making. For students who have earned more units on campus, the tree differentiates further between students based on their SAT composite score. Students with a score up to 990 are recommended to enroll in the statistics recitation class and attend MSLC at least once (node 2). The remaining students in node 3 are recommended to enroll also in the statistics recitation class but attend SI at least once instead of attending the MSLC. One possible interpretation could be that the MSLC is one-on-one tutoring which can therefore address the needs of weaker students more efficiently than SI.

Looking at the distribution of characteristics in each node, one notices an interesting pattern. Node 1 and node 3 gather students whose distributions of characteristics do not deviate strongly from the overall student sample displayed in Tables 4 and 5. Yet, node 2 collects students who are predominantly female (75% in node 2 vs. 60% in the overall data) from an underrepresented minority (63% in node 2 vs. 40% in

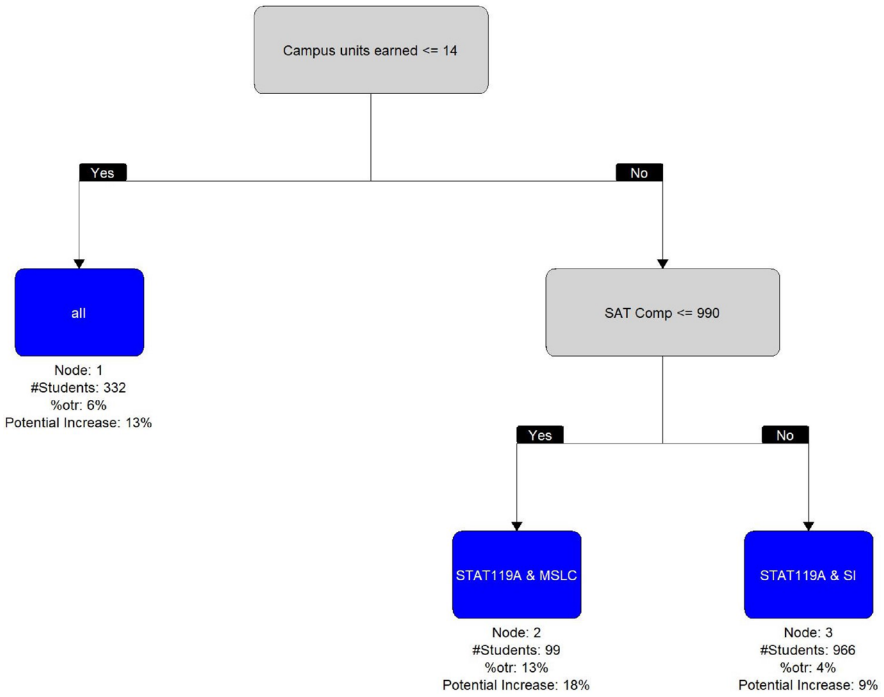


**Fig. 3** Tree grown with the help of the purity measure $\mathcal{P}^{\text{Tao}}$

the overall data). Also, the share of students who are the first generation at college is disproportionately high (35% in node 2 vs. 23% in the overall data). In terms of performance, the students in node 2 are weaker introductory statistics students with an average grade of 683 points compared to 776 points overall (one-sided $t$ test $p$ value of $1.78 \times 10^{-5}$). The share of students with a total GPA of $A$ in the node even indicates that those students are weaker in their studies in general (7.1% in node 2 vs. 32% in the overall data).

The described OTR was actually chosen by only 5% of the students who are distributed over all three nodes. In general, those students tend to perform well in their overall studies (92% of them have a Total GPA of at least an $B$). A bias-corrected and accelerated (BCa) bootstrap 95% interval of [53,166] of the difference in average grades suggests that choosing the treatment optimally could increase the average grade (based on 10,000 bootstrap samples). In our case, the average grade could improve from 776 to 856 points, i.e., from the letter grade $C$ to $B$.

Applying the purity measure $\mathcal{P}^{LZ}$ defined in (4) yields a tree with three instead of two splits, as shown in Fig. 4. According to this tree, weaker high school students, i.e., students with a high school GPA of at most 3.69, should enroll in the statistics recitation class and attend SI at least once (node 1). One possible explanation is that students who showed weaker performance during high school might struggle with learning at university, which requires potentially more independent study, and should therefore attend the statistics recitation class and its weekly

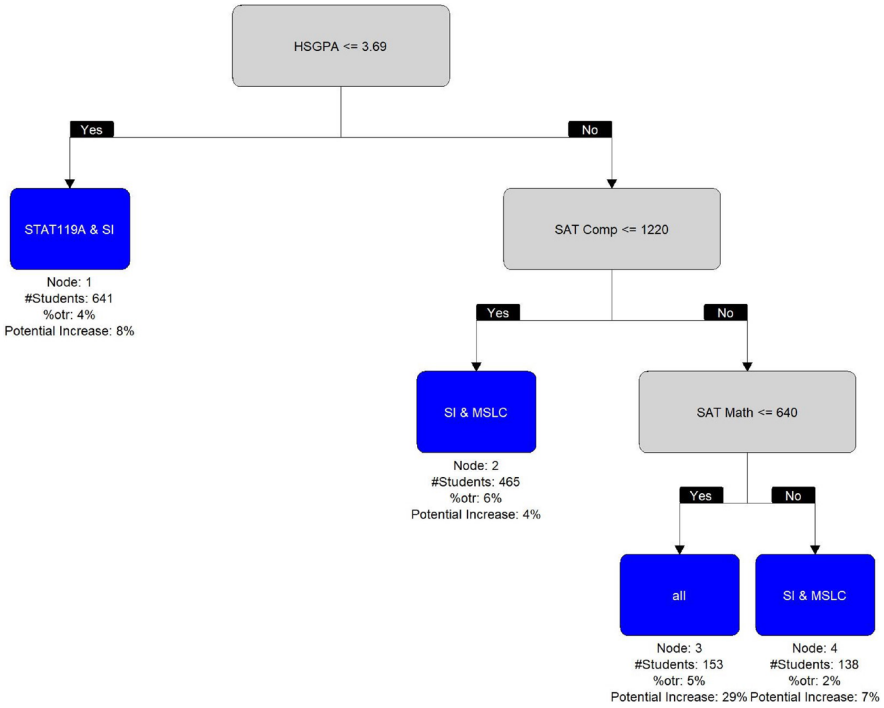


**Fig. 4** Tree grown with the help of the purity measure $\mathcal{P}^{LZ}$

lectures and assignments. For the better performing high school students, the tree recommends attendance in at least one session of SI and MSLC (nodes 2 and 3), *except* for students with an SAT composite score of above 1220 and an SAT math score of at most 640 (node 4). Those students are additionally recommended to enroll in the statistics recitation class. While it is intuitively reasonable to provide students with high SAT composite score but low SAT math score with additional training, it is not immediately clear why this recommendation shall not be given to students with a low SAT composite score as a poor score might stem from a poor SAT math score. However, the potential increase in the average grade of 29% for students with high SAT composite score and low SAT math score justifies the recommendation of all three programs.

Looking at the distributions of the characteristics, nodes 3 and 4 are striking. They exhibit a low share of people with an underrepresented minority background (27% in node 3 and 18% in node 4 against 40% in the overall data). Furthermore, the students are less frequently the first generation at college (14% in node 3 and 12% in node 4 against 23% in the overall data). The students perform well in the introductory statistics course with an average grade of 856 points in node 3 and 884 points in node 4 compared to an overall average grade of 776 points (one-sided $t$ test $p$ value smaller than $2.2 \cdot 10^{-16}$ in both cases). The students also have a higher

**Table 6** Comparison between *HSGPA* and *Campus units earned* at the first split

| Variable | $\mathcal{P}^{\text{Tao}}$ | $\mathcal{P}^{LZ}$ |
|---|---|---|
| Campus units earned | **849** (**1st**) | 830 (7th) |
| HSGPA | 844 (3th) | **843** (**1st**) |

share of Total GPAs of *A* (53% in node 3 and 69 % in node 4 against 32% in the overall data).

Again, a 95% BCa interval of [42,157] of the difference in average grades suggests that choosing the treatment optimally could increase the average grade (based on 10 000 bootstrap samples). In our case, the average grade could improve from 776 to 844 points, i.e., from the letter grade *C* to *B*. Only 4.5% chose their treatment optimally in the sense of Definition 1. Again, those students are present in all nodes and tend to be academically stronger in their overall studies (89% of them have a total GPA of at least an *B*).

Overall, the trees agree for 30% of the students (414 in total) in their treatment recommendation. Comparing both trees, the two promise to potentially increase the average grade significantly if the suggested treatment regime is implemented. Students seem to need a recommendation in the first place, as only 5% or 4.5%, respectively, of the students chose their treatment optimally in these semesters of introductory statistics. Yet, the trees differ significantly from each other in terms of the balance of their recommendation. Applied to our dataset, the purity measure $\mathcal{P}^{\text{Tao}}$ is proposing to split the students into one large group of about 70% of all students and two small groups with less than 10% and 20% of all students, respectively. The split recommended by the purity measure $\mathcal{P}^{LZ}$ is in this regard more balanced with the first node splitting the data almost in two halves. Regarding the choice of variables, it is striking that both trees use SAT scores to assign treatments. However, for the first split, $\mathcal{P}^{\text{Tao}}$ suggests as the split variable *Campus units earned*, whereas $\mathcal{P}^{LZ}$ proposes *HSGPA*. Table 6 shows that the use of either of the variables is considered reasonable by both purity measures as *Campus units earned* yields the seventh highest purity for $\mathcal{P}^{LZ}$ and *HSGPA* yields the third highest purity for $\mathcal{P}^{\text{Tao}}$.

The most important difference between the trees lies in the interpretability. As mentioned before, the tree grown with $\mathcal{P}^{LZ}$ exhibits an inconsistency at first sight when recommending the statistics recitation class to students with high SAT composite scores and low SAT math scores, but not to students with low SAT composite scores and low SAT math scores; even though the high potential increase in the average grade in the first group justifies it at a second glance. This apparent inconsistency might raise concern among practitioners and make them reluctant to implement the recommended treatment regime. Therefore, we propose to use the more consistent tree grown with $\mathcal{P}^{\text{Tao}}$. The fact that this tree also yields a higher potential average grade $\mathbb{E}[Y^*\{\pi(X)\}]$ makes this choice consistent with the definition of the OTR in (1).

## 5 Discussion and conclusions

Originated in the personalized medicine literature, this paper applies the concept of *optimal treatment regimes*—to our knowledge for the first time—to an educational data mining problem. We used the tree-based reinforcement learning approaches by Laber and Zhao (2015) and Tao et al. (2018). Other methods to estimate the OTR as random forests or a reinforcement learning approach based on neural networks, or alternative ensemble learning procedures are interesting to explore. However, leveraging those methods would most likely be at the expense of the interpretability of results and thus potentially less desirable to key university administrative stakeholders in the recommendation process.

Our estimated OTR gives students a recommendation based on their amount of campus units already earned and their SAT composite score. We estimate that this OTR potentially increases the average letter grade from an *C* to an *B*, indicating that the implementation of the OTR is likely to foster student success in the introductory statistics course. Furthermore, it became evident that students exhibit a need for an individual recommendation as for the given dataset only roughly 5% of students chose their treatment optimally.

Interestingly, no students are recommended to not attend any of the student success programs. This is a double-edged-sword suggesting on one hand that the statistics recitation class, MSLC, and SI program do indeed foster student success for every student. On the other hand, the implementation of the OTR becomes potentially infeasible as it would lead to a significant increase in resources from its current implementation. For instance, 210 students out of the roughly 1000 introductory statistics students in the Spring 2020 semester enrolled in the statistics recitation class. Implementing the OTR to the letter would therefore mean that SDSU needs to increase the capacity of the statistics recitation class by a factor of five. One cost efficient way to achieve this is to allow more students to the existing classes. However, increasing capacity in such a way might lead to a violation of Assumption 1. For example, significantly more students in each class would almost certainly impact the program's quality negatively, to wit: the treatment of one student affects other students' potential outcome, i.e., there are spill-over effects. Therefore, incorporating resource constraints in the estimation procedure is not only important from the educational institution's financial point of view but also from a theoretical one.

The personalized medicine literature approaches the resource allocation problem as a trade-off between health gains, side effects, and intervention costs. For example, Xu et al. (2020) presents an OTR to incorporate cost-effectiveness into the individualized treatment decision. On a front perhaps more in line with resource allocation problems in education settings, Luedtke and van der Laan (2016) and Toth and van der Laan (2018) include cost constraints in estimating the OTR for binary treatment. Both approaches are quite involved requiring additional constraints in setting up the optimization routines, the latter two papers being mathematical expositions on the efficacy of the proposed method. As significant algorithm and software development are required to bring these methods to fruition in our educational data mining context, we shall pursue these ideas in a future research study.

Though these resource allocation approaches are beyond the scope of this paper, a pilot subgroup analysis of our data may hint at the benefit of a cost-constrained OTR. The results in Sect. 4 recommend that every student shall attend at least two programs. If we reduce the data set to only those students that attended exactly one program, we may explore the trade-off between the programs. We term this a preliminary study for discussion purposes as statistics students were not limited to a choice of only one program and the programs were not designed as a "one-stop shop". This subgroup is thus a selective smaller sample of 503 students. The tree grown with the $\mathcal{P}^{\text{Tao}}$ purity measure recommends all 503 students attend SI, presenting a 2% potential increase in performance. The tree grown with the $\mathcal{P}^{LZ}$ purity measure recommends students with a high school GPA below 3.48 attend the MSLC (110 students; 6% potential increase in performance), and students above an HSGPA of 3.48 attend SI (393 students; 3% potential increase in performance). We note that the potential increases are not as strong as the analyses in Sect. 4 and neither approach recommends the Statistics recitation class as an option.

A reviewer brought up a query of whether any of these three programs may lower a student's course grade. We believe that negative treatment effects are possible. For example, stronger students who attend these programs may substitute attendance for study time and consequently perform worse on assessments. Or students who attend these programs may be less motivated students that are using attendance to avoid the harder work of studying. Alternatively, students who attend these programs may find the specific material harder (in that week or in general), thus motivating them to attend. And there could be a diminishing return along these lines by attending multiple programs. We note also that if we believe it is impossible for a program attendee to have lower course performance, then the strong ignorability assumption does not hold. Nonetheless, as the reviewer mentions, bias from unmeasured confounding or sampling error could result in the trees recommending a student into one or two programs instead of all three. To this end, the pilot analysis restricting to a subgroup of students that chose to attend exactly one program is free of these potential unmeasured biases. As we collect further semesters of data (larger sample size), these analyses along with a formal methodology of OTR under cost constraints may be of great use for program redesigns, student advising, and resource allocation.

Our analysis took only the spring semesters of 2018 and 2019 into account as the MSLC data set was unfortunately not available for the fall semester of 2018 and incomplete for the fall semester of 2019. Our future analyses will consider whether fall semesters, with their differently structured student body of usually more first-time-freshmen than in spring semesters, heavily impact the OTR.

Future research could bring our analysis to the next level by including the students' performance in, for instance, homeworks throughout the semester as demonstrated by Meier et al. (2016) in the context of grade prediction. This would eventually lead to the estimation of a *dynamic treatment regime* which would, for example, allow one to change the treatment recommendation weekly according to the ongoing performance of a student throughout a semester. For instance, the university could recommend a student to attend an SI session in week 5 after a poor performance on an assessment in week 4. A possible approach for the purity measure $\mathcal{P}^{\text{Tao}}$ is laid out in Tao et al. (2018).

In addition, future research could take societal values and legal restrictions regarding the variable choice into account. On one hand, including sensitive variables such as underrepresented minority status or gender in the estimation procedure might induce discrimination, although they do not appear in our trees. On the other hand, omitting those variables might still lead to decisions which are *counterfactually unfair* as Kusner et al. (2017) argues. The approach of Kusner et al. (2019) to impose constraints on *counterfactual privilege* to reduce the discriminatory impact of sensitive variables is promising in this regard.

Finally, the application of OTRs in educational data mining is not limited merely to recommendations regarding the student's optimal course choice. In fact, the method may be applied to any set of treatments and any student success outcome such as student retention, student probation rates, or graduation rates. Like this, OTRs might become relevant for college-prep programs or even scholarship programs under the condition that the mentioned Assumptions 1 to 4 are met. Note that Assumption 4—demanding access to each treatment for every student—may prove to be restrictive in this context.

## Proofs

**Lemma 1** 1 *If Assumptions* 1 *through* 4 *hold, we have*

$$\pi^{\text{opt}}(X) = \arg\max_{\pi \in \Pi} \mathbb{E}\left[\frac{Y \mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right].$$

***Proof*** Recall the definition of the OTR in Eq. (1)

$$\pi^{\text{opt}} = \arg\max_{\pi \in \Pi} \mathbb{E}[Y^*\{\pi(X)\}].$$

Hence, it is sufficient to show that

$$\mathbb{E}[Y^*\{\pi(X)\}] = \mathbb{E}\left[\frac{Y \mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right].$$

We begin using the law of iterated expectation to get

$$\begin{aligned} \mathbb{E}[Y^*\{\pi(X)\}] &= \mathbb{E}[\mathbb{E}\{Y^*\{\pi(X)\}|X]] \\ &= \mathbb{E}\left[\mathbb{E}\left[\sum_{a \in \mathcal{A}} Y^*(a) \mathbb{1}_{\{\pi(X)=a\}} \middle| X\right]\right]. \end{aligned} \tag{13}$$

Now, we leverage Assumption 3, stating that—conditioned on the characteristics $X$—all potential outcomes $Y^*(a)$ are independent of the treatment $A$ to add the condition in (13) that we assign treatment according to a treatment regime $\pi(X)$ which gives us

$$\begin{aligned}\mathbb{E}[Y^*\{\pi(X)\}] &= \mathbb{E}\left[\mathbb{E}\left[\sum_{a\in\mathcal{A}} Y^*(a)\mathbb{1}_{\{\pi(X)=a\}}|X, A=\pi(X)\right]\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}\left[\sum_{a\in\mathcal{A}} Y^*(a)\mathbb{1}_{\{\pi(X)=a\}}\mathbb{1}_{\{A=\pi(X)\}}|X\right]}{P\{A=\pi(X)|X\}}\right],\end{aligned} \tag{14}$$

where we used in the last step the definition of the conditional expectation. Note that the expression in Eq. (14) is well defined as Assumption 4 guarantees $P\{A=\pi(X)|X\}>0$. As

$$\begin{aligned}\mathbb{1}_{\{\pi(X)=a\}}\mathbb{1}_{\{A=\pi(X)\}} &= \begin{cases} 1 & \text{if } \pi(X)=a \text{ and } A=\pi(X) \\ 0 & \text{otherwise}\end{cases} \\ &= \begin{cases} 1 & \text{if } \pi(X)=a=A \\ 0 & \text{otherwise}\end{cases} \\ &= \begin{cases} 1 & \text{if } A=a \text{ and } A=\pi(X) \\ 0 & \text{otherwise}\end{cases} \\ &= \mathbb{1}_{\{A=a\}}\mathbb{1}_{\{A=\pi(X)\}},\end{aligned} \tag{15}$$

we can rewrite Eq. (14) as

$$\begin{aligned}\mathbb{E}[Y^*\{\pi(X)\}] &= \mathbb{E}\left[\frac{\mathbb{E}\left[\sum_{a\in\mathcal{A}} Y^*(a)\mathbb{1}_{\{A=a\}}\mathbb{1}_{\{A=\pi(X)\}}|X\right]}{P\{A=\pi(X)|X\}}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}\left[\left\{\sum_{a\in\mathcal{A}} Y^*(a)\mathbb{1}_{\{A=a\}}\right\}\mathbb{1}_{\{A=\pi(X)\}}|X\right]}{P\{A=\pi(X)|X\}}\right] \\ &= \mathbb{E}\left[\frac{\mathbb{E}\left[Y\mathbb{1}_{\{A=\pi(X)\}}|X\right]}{P\{A=\pi(X)|X\}}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A=\pi(X)|X\}}\middle|X\right]\right] \\ &= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A=\pi(X)|X\}}\right],\end{aligned} \tag{16}$$

where we used Assumption 2 in Eq. (16) and that $P\{A=\pi(X)|X\}$ is a $\sigma(X)$-measurable function, where $\sigma(X)$ is the $\sigma$-algebra generated by $X$. This ends our proof. □

**Lemma 2** *If Assumptions 1 through 4 hold, we have*

$$\pi^{\text{opt}}(X) = \arg\max_{\pi\in\Pi}\mathbb{E}\left[\frac{\{Y-g(X)\}\mathbb{1}_{\{A=\pi(X)\}}}{P\{A=\pi(X)|X\}}\right].$$

*for any arbitrary function* $g : \mathbb{R}^p \mapsto \mathbb{R}$ (*Laber and Zhao* 2015).

***Proof*** Let $g : \mathbb{R}^p \mapsto \mathbb{R}$ be an arbitrary function to define

$$L_g\{\pi(X)\} = \frac{\{Y - g(X)\}\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}. \tag{17}$$

Then, it holds that

$$\begin{aligned}
\mathbb{E}[L_g\{\pi(X)\}] &= \mathbb{E}\left[\frac{\{Y - g(X)\}\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] \\
&= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\left[\frac{g(X)\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] \\
&= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\left[\mathbb{E}\left[\frac{g(X)\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\middle|X\right]\right] \\
&= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\left[\frac{g(X)}{P\{A = \pi(X)|X\}}\mathbb{E}[\mathbb{1}_{\{A=\pi(X)\}}|X]\right] \\
&= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\left[\frac{g(X)}{P\{A = \pi(X)|X\}}P\{A = \pi(X)|X\}\right] \\
&= \mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\big[g(X)\big],
\end{aligned}$$

and therefore

$$\begin{aligned}
\arg\max_{\pi\in\Pi}\mathbb{E}[L_g\{\pi(X)\}] &= \arg\max_{\pi\in\Pi}\mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] - \mathbb{E}\big[g(X)\big] \\
&= \arg\max_{\pi\in\Pi}\mathbb{E}\left[\frac{Y\mathbb{1}_{\{A=\pi(X)\}}}{P\{A = \pi(X)|X\}}\right] \\
&= \pi^{\text{opt}}(X).
\end{aligned}$$

□

## Adjusted R code of Tao et al. (2018)

### Function `DTRtree` to grow tree with $\mathcal{P}^{\text{Tao}}$

```
DTRtree<-function(Y,A,X,m.method = c("LinearRegression","RandomForest"),ps.hat=
    NULL,mus.reg=NULL,depth=5,lambda.pct=0.05,minsplit=20){
  # initialization
  n<-length(Y)
  A<-as.factor(A)
  X<-as.matrix(X)
  I.node<-rep(1,n)
  class.A<-sort(unique(A))
  output<-matrix(NA,1,5)
  colnames(output)<-c("node","X","cutoff","Purity.Tao","trt")
  #Step 2: Estimate Propensity Scores and Conditional Mean Outcomes (if not
      given) using entire data
  if(is.null(ps.hat)) ps.hat<-M.propen(A=A,Xs=X) #multinomial regression
  if(is.null(mus.reg)==T){
    if(m.method=="LinearRegression"){
      KT<-length(unique(A))
      RegModel<-lm(Y ~ X*A)
      mus.reg<-matrix(NA,n,KT)
      for(k in 1L:KT) mus.reg[,k]<-predict(RegModel,newdata=data.frame(X,A=
          factor(rep(sort(unique(A))[k],n))))
    } else if(m.method=="RandomForest"){
      require(randomForest)
      RF<-randomForest(Y~., data=data.frame(A,X))
      mus.reg<-matrix(NA,n,length(class.A))
      for(i in 1L:length(class.A)){
        newdata=data.frame(X,A=as.factor(rep(class.A[i],n)))
        levels(newdata$A) <- levels(A)
        mus.reg[,i]<-predict(RF,newdata)
      }
    }
  }
  #Step 3: Compute the AIPW-estimator (TAO)
  mus.hat<-mus.AIPW(Y=Y,A=A,ps.hat=ps.hat,mus.reg=mus.reg)
  #Step 4: Split if purity gain is sufficient
  # expected outcome at root node
  root<-Opt.A(A,mus.hat)
  Ey0<-root$Ey.opt1
  # split if improved at least lambda, as a percent of Ey0
   for(k in 1L:depth){
    output<-rbind(output,matrix(NA,2^k,5))
    output[,1]<-1L:(2^(k+1)-1)
    if(k==1L){ #first layer of tree
      #find best split
      best.H.1<-best.H(H=X,A=A,mus.hat=mus.hat,minsplit=minsplit)
      #check purity gain requirement
      if(is.null(best.H.1)==F && best.H.1$mEy.opt1>Ey0*(1+lambda.pct)){
        #sufficient purity gain
        #left node: store in output row 2*k
        output[k,-1]<-c(best.H.1$X, best.H.1$X.subset, best.H.1$mEy.opt1, NA)
        I.node[I.node==k & X[,best.H.1$X] <= best.H.1$X.subset]<-2*k
        output[2*k,-1]<-c(NA,NA,NA,best.H.1$trt.L)
        #rightnode: store in output row 2*k+1
        I.node[I.node==k & X[,best.H.1$X] > best.H.1$X.subset]<-2*k+1
        output[2*k+1,-1]<-c(NA,NA,NA,best.H.1$trt.R)
      } else{ #not sufficient purity gain, choose single best treatment
        output[k,4:5]<-c(root$Ey.opt1,root$trt.opt1)
        break
      }
    }else{ #Rest of Layers of the tree
      for(j in (2^(k-1)):(2^k-1)){
        if(!is.na(output[trunc(j/2),2])){
          #find best split
```

```
        best.H.j<-best.H(H=X[I.node==j,],A=A[I.node==j],mus.hat=mus.hat[I.node
            ==j,],minsplit=minsplit)
        #check purity gain requirement
        if(is.null(best.H.j)==F && best.H.j$mEy.opt1>output[trunc(j/2),4]*(1+
            lambda.pct)){
          #sufficient purity gain -> split
          #left node: store in output row 2*j
          output[j,-1]<-c(best.H.j$X, best.H.j$X.subset, best.H.j$mEy.opt1, NA
              )
          I.node[I.node==j & X[,best.H.j$X] <= best.H.j$X.subset]<-2*j
          #right node: store in output row 2*j+1
          output[2*j,-1]<-c(NA,NA,NA,best.H.j$trt.L)
          I.node[I.node==j & X[,best.H.j$X] > best.H.j$X.subset]<-2*j+1
          output[2*j+1,-1]<-c(NA,NA,NA,best.H.j$trt.R)
        }
      }
    }
    if(sum(is.na(output[(2^(k-1)):(2^k-1),2]))==2^(k-1)) break
  }
 }
 output<-output[!is.na(output[,2]) | !is.na(output[,5]),]
 return(output)
}
```

**Function** `LZtree` **to grow tree with** $\mathcal{P}^{LZ}$

```
LZtree<-function(Y,A,X,m.method=c("LinearRegression","RandomForest"),ps.hat=
    NULL,mus.reg=NULL,depth=5,lambda.pct=0.05,minsplit=20){
  # initialization
  n<-length(Y)
  I.node<-rep(1,n)
  A<-as.factor(A)
  X<-as.matrix(X)
  class.A<-sort(unique(A))
  output<-matrix(NA,1,5)
  colnames(output)<-c("node","X","cutoff","Purity.LZ","trt")
  #Step 2: Estimate Propensity Scores and Conditional Mean Outcomes (if not
      given) using entire data
  if(is.null(ps.hat))ps.hat<-M.propen(A=A,Xs=X)
  if(is.null(mus.reg)==T){
    if(m.method=="LinearRegression"){
      KT<-length(unique(A))
      RegModel<-lm(Y ~ X*A)
      mus.reg<-matrix(NA,n,KT)
      for(k in 1L:KT) mus.reg[,k]<-predict(RegModel,newdata=data.frame(X,A=
          factor(rep(class.A[k],n))))
    } else if(m.method=="RandomForest"){
      require(randomForest)
      RF<-randomForest(Y~., data=data.frame(X,A))
      mus.reg<-matrix(NA,n,length(class.A))
      for(i in 1L:length(class.A)){
        newdata=data.frame(H,A=as.factor(rep(class.A[i],n)))
        levels(newdata$A) <- levels(A)
        mus.reg[,i]<-predict(RF,newdata)
      }
    }
  }
  #Step 3: Compute the estimator for m(x) (LABER)
  mus.max<-apply(mus.reg,1,max)
  #Step 4: Split if the purity gain is sufficient
  # expected outcome at root node
  ps.A<-rep(NA,n)
  for(i in 1:n){
    ps.A[i]<-ps.hat[i,which(class.A==A[i])]
  }
  root<-Opt.A.LZ(Y,A,mus.max,ps.A)
  Ey0<-root$Ey.opt1
  for(k in 1L:depth){
    output<-rbind(output,matrix(NA,2^k,5))
    output[,1]<-1L:(2^(k+1)-1)
    if(k==1L){#first layer of tree
      #find best split
      best.H.1<-best.H.LZ(H=X,Y=Y,A=A,mus.max=mus.max,ps.A=ps.A,minsplit=
          minsplit)
      #check purity gain requirement
      if(is.null(best.H.1)==F && best.H.1$mEy.opt1>Ey0*(1+lambda.pct)){
        #sufficient purity gain
        output[k,-1]<-c(best.H.1$X, best.H.1$X.subset, best.H.1$mEy.opt1, NA)
        #left node: store in output row 2*k
        I.node[I.node==k & X[,best.H.1$X] <= best.H.1$X.subset]<-2*k
        output[2*k,-1]<-c(NA,NA,NA,best.H.1$trt.L)
        #right node: store in output row 2*k+1
        I.node[I.node==k & X[,best.H.1$X] > best.H.1$X.subset]<-2*k+1
        output[2*k+1,-1]<-c(NA,NA,NA,best.H.1$trt.R)
      } else{#not sufficient purity gain, choose single best treatment
        output[k,4:5]<-c(root$Ey.opt1,root$trt.opt1)
        break
      }
```

```
    } else{#Rest of Layers of the tree
      for(j in (2^(k-1)):(2^k-1)){
        if(!is.na(output[trunc(j/2),2])){
          #find best split
          best.H.j<-best.H.LZ(H=X[I.node==j,],Y=Y[I.node==j],A=A[I.node==j],mus.
               max=mus.max[I.node==j],
                           ps.A=ps.A[I.node==j],minsplit=minsplit)
          #check purity gain requirement
          if(is.null(best.H.j)==F && best.H.j$mEy.opt1>output[trunc(j/2),4]+
               lambda){
            #sufficient purity gain requirement
            #left node: store in output row 2*j
            output[j,-1]<-c(best.H.j$X, best.H.j$X.subset, best.H.j$mEy.opt1, NA
                 )
            I.node[I.node==j & X[,best.H.j$X] <= best.H.j$X.subset]<-2*j
            output[2*j,-1]<-c(NA,NA,NA,best.H.j$trt.L)
            #right node: store in output row 2*j+1
            I.node[I.node==j & X[,best.H.j$X] > best.H.j$X.subset]<-2*j+1
            output[2*j+1,-1]<-c(NA,NA,NA,best.H.j$trt.R)
          }
        }
      }
      if(sum(is.na(output[(2^(k-1)):(2^k-1),2]))==2^(k-1)) break
    }
  }
  output<-output[!is.na(output[,2]) | !is.na(output[,5]),]
  return(output)
}
```

### Example

This subsection provides the R-code of the file `03 Example Application` to demonstrate how to deploy the developed methods. The functions `DTRtree` and `LZtree` are available in the file `01 TRL Functions`—along with other functions.

```
####Required Packages####
#none
####Required Functions####
source("01 TRL Functions.R")
```

The example considers simulated data in the file `02 Example Data` which presents a similar structure to the student success data, but is not based on actual student data. For each student, an SAT Math Score, HSGPA, Age, Gender, and URM were simulated along with an overall grade. Three treatments were assumed to be available to every student: no program (encoded with 1), MSLC (encoded with 2), and SI (encoded with 3).

```
####Dataset####
data<-read.csv("02 Example Data.csv",fileEncoding="UTF-8-BOM")
data.Y<-data$Y #observed outcome
data.X<-data[1:5] #characteristics
data.A<-data$A #observed treatment
```

To estimate an OTR for the given data either using $\mathcal{P}^{\text{Tao}}$ or $\mathcal{P}^{LZ}$, we need to choose at first a maximal tree depth, a minimal purity gain $\lambda$, and a minimal node size $\gamma$ following our algorithm in Sect. 2.3.

```
####Parameters####
depth<-5 #maximal tree depth
lambda<-0.025 #minimal purity gain
gamma<-max(nrow(data)*0.05,20) #minimal node size
```

The remaining steps of the algorithm are then performed by the functions `DTRtree` for $\mathcal{P}^{\text{Tao}}$ and `LZtree` for $\mathcal{P}^{LZ}$.

```
####TaoTree####
taotree<-DTRtree(Y=data.Y,A=data.A,X=data.X,m.method="LinearRegression",lambda.
     pct=lambda,minsplit=gamma,depth=depth)
####LaberTree####
labertree<-LZtree(Y=data.Y,A=data.A,X=data.X,m.method = "LinearRegression",
     lambda.pct=lambda,minsplit=gamma,depth = depth)
```

The output `taotree` is given as a matrix:

```
> taotree
     node X cutoff Purity.Tao trt
[1,] 1 3 18 607.5887 NA
[2,] 2  NA NA NA 2
[3,] 3 1 580 644.4112 NA
[4,] 6  NA NA NA 3
[5,] 7  NA NA NA 1
```

For example, the first node is split with the help of the third covariate—which is age—at a value of 18. The purity $\mathcal{P}^{\text{Tao}}$ associated with this split is 608. No treatment is assigned as node 1 is not a terminal node. All students who are at most 18 are sent to node 2 which assigns treatment 2 (MSLC) to each student. All students who are older than 18 are differentiated according to their SAT Math score—which is the first covariate in the dataset. If they achieved an SAT Math Score of at most 580, they are recommended to attend SI (encoded as 3); otherwise, they should not attend any program (encoded as 1). Figure 5 displays the graphical representation of the output `taotree`.

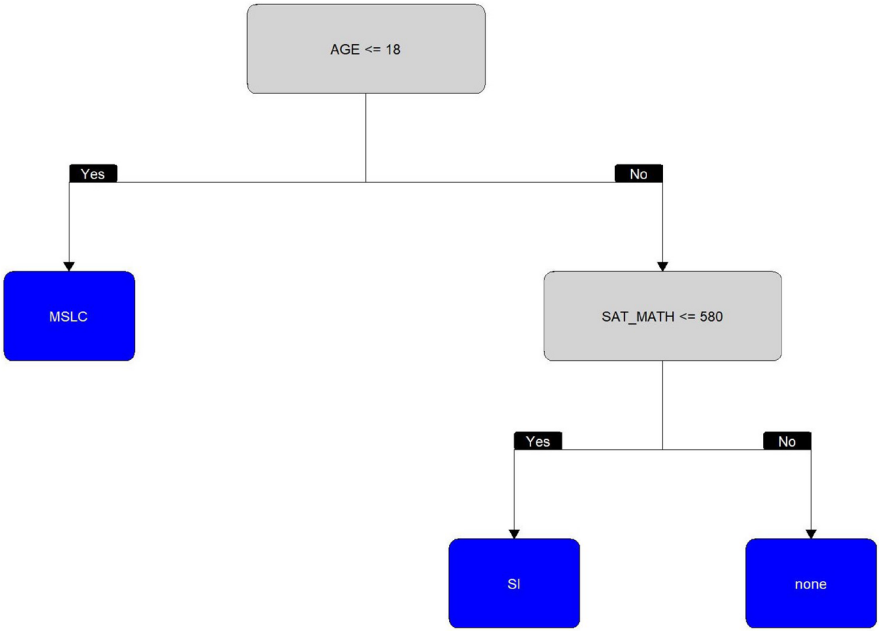


**Fig. 5** Example tree grown with the help of the purity measure $\mathcal{P}^{\text{Tao}}$

## References

Bang H, Robins JM (2005) Doubly robust estimation in missing data and causal inference models. Biometrics 61(4):962–973

Chernozhukov V, Chetverikov D, Demirer M, Duflo E, Hansen C, Newey W, Robins J (2018) Double debiased machine learning for treatment and structural parameters. Economet J 21(1):C1–C68

Doubleday K, Zhou H, Fu H, Zhou J (2018) An algorithm for generating individualized treatment decision trees and random forests. J Comput Graph Stat 27:849–860

Gill RD, Robins JM (2001) Causal inference for complex longitudinal data: the continuous case. Ann Stat 29(6):1785–1811

Kusner M, Russell C, Loftus J, Silva R (2019) Making decisions that reduce discriminatory impacts. In: Chaudhuri K, Salakhutdinov R (eds) Proceedings of the 36th international conference on machine learning, PMLR, proceedings of machine learning research, vol 97, pp 3591–3600

Kusner MJ, Loftus J, Russell C, Silva R (2017) Counterfactual fairness. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (eds) Advances in neural information processing systems, vol 30. Curran Associates Inc., New York, pp 4066–4076

Laber EB, Zhao YQ (2015) Tree-based methods for individualized treatment regimes. Biometrika 102:501–514

Luedtke AR, van der Laan MJ (2016) Optimal individualized treatments in resource-limited settings. Int J Biostat 12(1):283–303

Martin DC, Arendale DR (1992) Supplemental instruction: improving first-year student success in high-risk courses (2nd ed). National Resource Center for The First Year Experience

Meier Y, Xu J, Atan O, van der Schaar M (2016) Predicting grades. IEEE Trans Signal Process 64(4):959–972

Powell MG, Hull DM, Beaujean AA (2020) Propensity score matching for education data: worked examples. J Exp Educ 88(1):145–164

Qian M, Murphy SA (2011) Performance guarantees for individualized treatment rules. Ann Stat 39(2):1180–1210

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/
Robins J (1986) A new approach to causal inference in mortality studies with a sustained exposure period-application to control of the healthy worker survivor effect. Math Model 7(9–12):1393–1512
Rosenbaum PR, Rubin DB (1983) The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55
Rubin DB (1980) Randomization analysis of experimental data: the fisher randomization test comment. J Am Stat Assoc 75(371):591–593
Schneider M, Preckel F (2017) Variables associated with achievement in higher education: a systematic review of meta-analyses. Psychol Bull 143:565–600
Sobel ME (2006) What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference. J Am Stat Assoc 101(476):1398–1407
Spoon MK, Beemer J, Whitmer J, Fan J, Frazee PJ, Andrew J, Bohonak JS, Levine AR (2016) Random forests for evaluating pedagogy and informing personalized learning. Educ Data Min 20:20–50
Tao Y, Wang L (2017) Adaptive contrast weighted learning for multi-stage multi-treatment decision-making. Biometrics 73(1):145–155
Tao Y, Wang L, Almirall D (2018) Tree-based reinforcement learning for estimating optimal dynamic treatment regimes. Ann Appl Stat 12(3):1914–1938
Toth B, van der Laan M (2018) Targeted learning of optimal individualized treatment rules under cost constraints. In: Biopharmaceutical applied statistics symposium, pp 1–22
VanderWeele TJ, Hernan MA (2013) Causal inference under multiple versions of treatment. J Causal Inference 1(1):1–20
Xu Y, Greene T, Bress A, Sauer B, Bellows B, Zhang Y, Weintraub W, Moran A, Shen J (2020) Estimating the optimal individualized treatment rule from a cost-effectiveness perspective. Biometrics 20:1–15
Zhang B, Tsiatis AA, Laber EB, Davidian M (2012) A robust method for estimating optimal treatment regimes. Biometrics 68:1010–1018
Zhao Y, Zeng D, Rush AJ, Kosorok MR (2012) Estimating individualized treatment rules using outcome weighted learning. J Am Stat Assoc 107(499):1106–1118