Published in final edited form as:

Biometrics. 2013 September ; 69(3): . doi:10.1111/biom.12052.

Inference for Optimal Dynamic Treatment Regimes using an Adaptive m -out-of- n Bootstrap Scheme

Bibhas Chakraborty^{1,*}, Eric B. Laber^{2,**}, and Yingqi Zhao^{3,***}

¹Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.

²Department of Statistics, North Carolina State University, Raleigh, NC, 27695, U.S.A.

³Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, 53792, U.S.A.

Summary

A dynamic treatment regime consists of a set of decision rules that dictate how to individualize treatment to patients based on available treatment and covariate history. A common method for estimating an optimal dynamic treatment regime from data is Q -learning which involves nonsmooth operations of the data. This nonsmoothness causes standard asymptotic approaches for inference like the bootstrap or Taylor series arguments to breakdown if applied without correction. Here, we consider the m -out-of- n bootstrap for constructing confidence intervals for the parameters indexing the optimal dynamic regime. We propose an adaptive choice of m and show that it produces asymptotically correct confidence sets under fixed alternatives. Furthermore, the proposed method has the advantage of being conceptually and computationally much more simple than competing methods possessing this same theoretical property. We provide an extensive simulation study to compare the proposed method with currently available inference procedures. The results suggest that the proposed method delivers nominal coverage while being less conservative than alternatives. The proposed methods are implemented in the `qLearn` R-package and have been made available on the Comprehensive R-Archive Network (<http://cran.r-project.org/>). Analysis of the Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study is used as an illustrative example.

Keywords

dynamic treatment regime; Q -learning; non-regularity; m -out-of- n bootstrap

1. Introduction

Dynamic treatment regimes (DTRs), also known as treatment policies, offer a framework for developing multistage personalized treatment strategies. A DTR consists of a sequence of decision rules, one per stage of clinical intervention, where each decision rule takes a patient's treatment and covariate history as input, and provides a recommended treatment as

*bc2425@columbia.edu

**eblaber@ncsu.edu

***yqzhao@biostat.wisc.edu

Supplementary Material

Web Appendix A, referenced in Sections 1 and 2, Web Appendix B, referenced in Sections 1 and 2, Web Appendix C, referenced in Sections 2 and 3, Web Appendix D, referenced in Section 2, and Web Appendix E, referenced in Section 4, are available with this paper at the Biometrics website on Wiley Online Library.

output. A DTR is said to be *optimal* if it yields the best clinical outcome on average when applied to the entire patient population.

Standard methods for estimating an optimal DTR are based on backwards recursive estimation procedures that mimic dynamic programming (Murphy, 2003; Robins, 2004). One such method that has gained popularity is *Q-learning* (Watkins, 1989; Sutton and Barto, 1998; Murphy, 2005b; Schulte et al., 2012); this method uses approximate dynamic programming with the requisite expectations being approximated with regression models. A number of variations on *Q-learning* have since been proposed (Zhao et al., 2009; Chakraborty et al., 2010; Shortreed et al., 2011; Song et al., 2012). However, inference for DTRs based on *Q-learning* and related approaches is nonstandard as the foregoing methods yield estimators with nonregular limiting distributions (Robins, 2004; Laber et al., 2012; Hirano and Porter, 2012). We focus on the problem of constructing valid confidence sets in a way that automatically adapts to nonregularity present in the underlying generative model. To be clear, we use the term nonregularity to refer to the phenomenon wherein abrupt changes in the limiting distribution of parameters used in *Q-learning* result from small perturbations in the underlying generative model. Our analysis is done under the paradigm of fixed alternatives.

The *m*-out-of-*n* bootstrap is a general-purpose tool for producing valid confidence sets for nonsmooth functionals (Bretagnolle, 1983; Swanepoel, 1986; Dümbgen, 1993; Shao, 1994; Shao and Wu, 1989; Huang et al., 1996; Bickel et al., 1997). By drawing $m < n$ samples with replacement from the original data, the *m*-out-of-*n* bootstrap acts as a smoothing operation on the empirical distribution function, with more smoothing corresponding to smaller values of the resample size *m* (we review the *m*-out-*n* bootstrap in Section 2.3). Intuitively, the choice of *m* should reflect a data-driven measure of nonregularity. Data-driven approaches for choosing *m* in other contexts were given by Hall et al. (1995), Lee (1999), Cheung et al. (2005) and Bickel and Sakov (2008). We apply the *m*-out-of-*n* bootstrap to construct confidence intervals (CIs) for parameters in *Q-learning* as well as *bias-reduced Q-learning* (*BRQ-learning*) (Chakraborty et al., 2010; Moodie and Richardson, 2010). We develop a data-driven, adaptive choice of *m* that is directly connected with a measure of nonregularity. We also extend the approach of selecting *m* in Bickel and Sakov (2008) to *Q-learning*.

We see the contributions of this paper as follows. We offer a novel method for choosing the resample size for nonregular estimators that is adaptive in that it leads to the usual *n*-out-of-*n* bootstrap when the generative model induces sufficient smoothness, and the *m*-out-of-*n* bootstrap otherwise. This methodology is developed for producing asymptotically valid confidence intervals for parameters indexing estimated optimal DTRs. The proposed methodology is conceptually and computationally simple, making it appealing for use in practice. To facilitate dissemination, we have developed an R package called *qLearn* implementing our method. This is freely available at the Comprehensive R Archive Network (CRAN): <http://cran.r-project.org/web/packages/qLearn/>.

In Section 2, we review *Q-learning*; propose our adaptive *m*-out-of-*n* resampling scheme; and extend an existing resampling method to *Q-learning*. Simulations comparing proposed methods with existing methods are presented in Section 3. We illustrate the proposed method using data from the STAR*D clinical trial on depression in Section 4. We conclude with a discussion in Section 5. Extension of the methodology to *BRQ-learning* is provided in Web Appendix A. Theoretical results and their proofs are deferred to Web Appendix B.

2. Set-up and Methodology

For simplicity, we focus on studies with two stages of clinical intervention. Longitudinal data on a single patient are given by the trajectory $(\mathbf{O}_1, A_1, \mathbf{O}_2, A_2, \mathbf{O}_3)$, where $\mathbf{O}_j$ ($j = 1, 2$) denotes the vector of covariates measured prior to treatment at the beginning of the j -th stage, $\mathbf{O}_3$ denotes the covariates measured at the end of stage 2, and A_j ($j = 1, 2$) denotes the treatment assigned at the j -th stage subsequent to observing $\mathbf{O}_j$. The data used to estimate an optimal DTR consists of a random sample of n independent and identically distributed (*i.i.d.*) trajectories. Define the history at each stage as: $\mathbf{H}_1 = \mathbf{O}_1, \mathbf{H}_2 = (\mathbf{O}_1^T, A_1, \mathbf{O}_1^T)^T$, thus, the history at stage j comprises the information available to the decision maker prior to the assignment of the j -th treatment. We consider studies in which there are two possible treatments at each stage, $A_j \in \{-1, 1\}$ for $j = 1, 2$, and treatment assignment is randomized (conditional on the history) with known randomization probabilities. This design is a special case of a sequential, multiple assignment, randomized trial (Lavori and Dawson, 2000, 2004; Dawson and Lavori, 2004; Murphy, 2005a). The study has two stage-specific outcomes Y_1 and Y_2 leading to an overall outcome $Y = Y_1 + Y_2$. We assume that $Y_1 = f_1(\mathbf{O}_1, A_1, \mathbf{O}_2)$ and $Y_2 = f_2(\mathbf{O}_1, A_1, \mathbf{O}_2, A_2, \mathbf{O}_3)$, for known functions f_1, f_2 . A two-stage DTR consists of a pair of decision rules, denoted by (d_1, d_2) , with d_j : $\text{domain}(\mathbf{H}_j) \rightarrow \{-1, 1\}$, so that d_j maps each patient history at stage j to one of the available treatments. The goal is to estimate the pair of decision rules that maximizes $\mathbb{E}_{d_1, d_2}(Y)$, where $\mathbb{E}_{d_1, d_2}$ denotes the expectation over the joint trajectory distribution under the restriction that $A_j = d_j(\mathbf{H}_j)$.

2.1 Basic Q-learning with Linear Models

One way to estimate the optimal DTR is via Q -learning (Watkins, 1989; Murphy, 2005b), which is a form of approximate dynamic programming. Define the Q -functions (Sutton and Barto, 1998; Murphy, 2005b) for the two stages as follows:

$$Q_2(\mathbf{H}_2, A_2) = E[Y_2 | \mathbf{H}_2, A_2], Q_1(\mathbf{H}_1, A_1) = E[Y_1 + \max_{a_2} Q_2(\mathbf{H}_2, a_2) | \mathbf{H}_1, A_1].$$

If the two Q -functions were known, then the decision rules comprising the optimal DTR can be found using dynamic programming (Bellman, 1957): $d_j^{DP}(\mathbf{h}_j) = \arg \max_{a_j} Q_j(\mathbf{h}_j, a_j)$, $j = 1, 2$. In practice, however, the true Q -functions are not known because the underlying generative distribution is not known, and hence they must be estimated from the observed data. We consider linear working models for the Q -functions of the form

$Q_j(\mathbf{H}_j, A_j; \beta_j, \psi_j) = \beta_j^T \mathbf{H}_{j0} + (\psi_j^T \mathbf{H}_{j1}) A_j$, where $\mathbf{H}_{j0}$ and $\mathbf{H}_{j1}$ are two (possibly different) summaries of the history $\mathbf{H}_j$. Both $\mathbf{H}_{j0}$ and $\mathbf{H}_{j1}$ include intercept terms. The Q -learning algorithm is as follows:

1. Stage-2 regression:

$$(\hat{\beta}_{2,n}, \hat{\psi}_{2,n}) = \arg \min_{\beta_2, \psi_2} \sum_{i=1}^n (Y_{2,i} - Q_2(\mathbf{H}_{2,i}, A_{2,i}; \beta_2, \psi_2))^2.$$

2. Stage-1 pseudo-outcome: $\hat{Y}_{1,i} = Y_{1,i} + \max_{a_2} Q_2(\mathbf{H}_{2,i}, a_2; \hat{\beta}_{2,n}, \hat{\psi}_{2,n})$, $i = 1, \dots, n$.

3. Stage-1 regression:

$$(\hat{\beta}_{1,n}, \hat{\psi}_{1,n}) = \arg \min_{\beta_1, \psi_1} \sum_{i=1}^n (\hat{Y}_{1,i} - Q_1(\mathbf{H}_{1,i}, A_{1,i}; \beta_1, \psi_1))^2.$$

The stage- j estimated optimal rule is given by $d_j^{\hat{}}(\mathbf{h}_j) = \arg \max_{a_j} Q_j(\mathbf{h}_j, a_j; \hat{\beta}_{j,n}, \hat{\psi}_{j,n})$, $j = 1, 2$, and thereby the estimated optimal DTR using Q -learning is given by $(d_1^{\hat{}}, d_2^{\hat{}})$. Thus, the Q -learning algorithm mimics the dynamic programming approach by plugging in fitted regressions for the requisite conditional expectations.

2.2 Nonregularity

It is now well-known that because of the nonsmooth maximization operation (step 2 above), Q -learning suffers from nonstandard limit theory (Robins, 2004). Note that the stage-1 pseudo-outcome in Q -learning for the i -th subject is

$$\hat{Y}_{1,i} = Y_{1,i} + \max_{a_2} Q_2(\mathbf{H}_{2,i}, a_2; \hat{\beta}_{2,n}, \hat{\psi}_{2,n}) = Y_{1,i} + \hat{\beta}_{2,n}^T \mathbf{H}_{20,i} + |\hat{\psi}_{2,n}^T \mathbf{H}_{21,i}|, i=1, \dots, n,$$

which is a nonsmooth (nondifferentiable at $\hat{\psi}_{2,n}^T \mathbf{H}_{21,i}=0$) function of $\hat{\psi}_{2,n}$. Let θ_1 denote the first stage parameters $(\beta_{1,n}^T, \psi_1^T)^T$, and $\hat{\theta}_{1,n}$ the corresponding vector of estimators based on a sample of size n . Since $\hat{\theta}_{1,n}$ is a function of $\hat{Y}_{1,i}$, $i=1, \dots, n$, it is in turn a nonsmooth function of $\hat{\psi}_{2,n}$. As a consequence, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{1,n} - \theta_1)$ does not converge uniformly over the parameter space of ψ_2 (Robins, 2004; Laber et al., 2012). More specifically, the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{1,n} - \theta_1)$ is normal if ψ_2 is such that $P[\mathbf{H}_{21} : \mathbf{H}_{21}^T \psi_2 = 0] = 0$, but is non-normal if $P[\mathbf{H}_{21} : \mathbf{H}_{21}^T \psi_2 = 0] > 0$. Thus, small perturbations in the marginal distribution of $\mathbf{H}_{21}$, the parameter ψ_2 , or both, can lead to abrupt differences in the limiting distribution of first stage estimators. We note that while the germane probability $P[\mathbf{H}_{21} : \mathbf{H}_{21}^T \psi_2 = 0]$ may indeed be zero, it is difficult in small samples to distinguish points that are near the boundary $\{\mathbf{h}_{21} : \mathbf{h}_{21}^T \psi_2 = 0\}$ from points that are actually on this boundary. Assuming that the foregoing probability can be positive is really a working assumption made to reflect the uncertainty about optimal treatment for patients with small treatment effects. This point has been developed and explored in Pötscher (2007), Pötscher and Schneider (2008), and Laber and Murphy (2011).

Owing to the aforementioned nonsmoothness, $\hat{\theta}_{1,n}$ is an asymptotically biased estimator of θ_1 , and Wald type confidence intervals for θ_1 show poor coverage rates (Robins, 2004; Moodie and Richardson, 2010). Furthermore, under certain simple generative models standard bootstrap confidence intervals also perform poorly (Chakraborty et al., 2010). BRQ -learning procedures based on thresholding were proposed by Chakraborty et al. (2010) and Moodie and Richardson (2010) (see Web Appendix A for a full description). While these procedures showed reductions in bias, and bootstrapping them yielded improved coverage when compared with Q -learning, they involve nonsmooth transformations of the data and theoretical justification is lacking. Recently Song et al. (2012) operationalized thresholding of the Q -learning estimator via a lasso-type penalization scheme.

2.3 The m -out-of- n Bootstrap

The m -out-of- n bootstrap is a general method for remedying bootstrap inconsistency due to nonsmoothness (see Shao, 1994; Bickel et al., 1997, and references therein; see also Politis et al., 1999). The m -out-of- n bootstrap is the usual nonparametric bootstrap (Efron, 1979) except that the resample size, historically denoted by m , is of a smaller order than the original sample size n . That is, m depends on n , tends to infinity with n , and satisfies $m = o(n)$. Intuitively, the m -out-of- n bootstrap works asymptotically by letting the empirical distribution tend to the true generative distribution at a faster rate than the analogous convergence of the bootstrap empirical distribution to the empirical distribution. In essence, this allows the empirical distribution to reach its limit 'first' so that bootstrap resamples behave as if they were drawn from the true generative distribution.

To assist with intuition and provide a brief review of the m -out-of- n bootstrap, we consider a toy example. Suppose $Z_1, \dots, Z_n$ are *i.i.d.* random variables with mean μ and unit variance,

and that we are interested in constructing a confidence interval for $\theta \triangleq |\mu|$ which is a nondifferentiable function of μ at zero. Let $\bar{Z}_n \triangleq n^{-1} \sum_{i=1}^n Z_i$ denote the sample mean, then $\hat{\theta}_n \triangleq |\bar{Z}_n|$ is a plug-in estimator of θ . Using the central limit theorem, it can be shown that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges in distribution to a standard normal random variable if $\mu \neq 0$, but converges in distribution to a χ_1 distribution (i.e., a half-normal distribution) if $\mu = 0$. Let $m = m(n)$ be an arbitrary sequence of positive integers and for any n let $W_{n1}, \dots, W_{nm}$ be Multinomial random variables with m trials and uniform success probability vector $(1/n, 1/n, \dots, 1/n)$. Define the bootstrap mean $\bar{Z}_m^{(b)} \triangleq m^{-1} \sum_{i=1}^m W_{ni} Z_i$. Provided m and n tend to ∞ , then conditional on $Z_1, \dots, Z_n$ (so that the randomness comes from the multinomial weights), it follows that $\sqrt{m}(\bar{Z}_m^{(b)} - \bar{Z}_n)$ converges in distribution to a standard normal random variable along almost all sample sequences $Z_1, Z_2, \dots$ (Bickel and Freedman, 1981, Theorem 2.1). Define $\hat{\theta}_m^{(b)} \triangleq |\bar{Z}_m^{(b)}|$; a bootstrap confidence interval is formed by estimating the quantiles of $\sqrt{m}(\hat{\theta}_m^{(b)} - \hat{\theta}_n)$, and using these to approximate the corresponding quantiles of $\sqrt{n}(\hat{\theta}_n - \theta)$. If $\mu \neq 0$ then it can be shown that $\sqrt{m}(\hat{\theta}_m^{(b)} - \hat{\theta}_n)$ converges in distribution to a standard normal random variable along almost all sample sequences provided n and m tend to ∞ (without restrictions on the rates of m and n , Bickel and Freedman, 1981). However, if $\mu = 0$, then

$$\sqrt{m}(\hat{\theta}_m^{(b)} - \hat{\theta}_n) = \left| \sqrt{m}(\bar{Z}_m^{(b)} - \bar{Z}_n) + \sqrt{\frac{m}{n}} \sqrt{n\bar{Z}_n} \right| - \left| \sqrt{\frac{m}{n}} \sqrt{n\bar{Z}_n} \right|, \quad (1)$$

which can be seen to converge conditionally on $Z_1, \dots, Z_n$ in distribution to a χ_1 random variable *only* when m tends to ∞ with n and satisfies $m = o(n)$. In particular, when $m = n$, which corresponds to the usual nonparametric bootstrap, (1) does not converge to a limit as $\sqrt{n}\bar{Z}_n$ does not converge (conditionally on $Z_1, \dots, Z_n$).

One difficulty associated with the m -out-of- n bootstrap is the choice of m since the condition $m = o(n)$ is purely asymptotic and thus provides no guidance for finite samples. Intuitively, the choice of the resample size m should reflect the degree of nonsmoothness in the underlying generative model. The nonsmoothness in Q -learning arises when there is an amassing of points near the boundary $\{\mathbf{h}_{21} : \mathbf{h}_{21}^T \psi_2 = 0\}$. Define $p \triangleq P(\mathbf{H}_{21}^T \psi_2 = 0)$. When $p = 0$, $\sqrt{n}(\hat{\theta}_{1,n} - \theta_1)$ is asymptotically normal and the n -out-of- n bootstrap is consistent. However, when $p > 0$, given that $\psi_{2,n}$ is not exactly equal to the true value, $\theta_{1,n}$, as a function of $\psi_{2,n}$, oscillates with a rate $n^{-1/2}$ around a point where abrupt changes of the asymptotic distribution occur. This is also true for its bootstrap analogue $\hat{\theta}_{1,m}^{(b)}$ while the oscillating rate is $m^{-1/2}$. With a large p , indicating a high degree of nonregularity, we hope this bootstrap analog oscillates with a rate much slower than $n^{-1/2}$. Therefore, we propose a data adaptive choice of m which directly reflects the underlying non-regularity measure. We also extend the method Bickel and Sakov (2008) for choosing resample sizes m to the context of Q -learning.

2.3.1 An Adaptive Scheme to Select m —We propose a class of resample sizes given by $m \triangleq n^{f(p)}$, where $f(p)$ is a function of p that satisfies: (C1) $f(p)$ is monotone decreasing in p , takes values in $(0, 1]$, and $f(0) = 1$; and (C2) $f(p)$ is continuous and has bounded first

derivative. Define the plug-in estimator for p , $\hat{p} = \mathbb{P}_n \mathcal{I} \{n[\mathbf{H}_{21}^T \hat{\psi}_{2,n}]^2 \leq \tau_n(\mathbf{H}_{21})\}$ (for cutoff $\tau_n(\mathbf{H}_{21})$, see below), where $\mathbb{P}_n$ denotes the empirical measure; subsequently define $m \triangleq n^{\hat{p}}$.

In Web Appendix B, we prove that $m/\hat{n}^{f(p)} \rightarrow 1$ almost surely, and thus $\hat{m} \xrightarrow{p} \infty$ and $\hat{m}/n \xrightarrow{p} 0$. For implementation, we propose a simple form of $f(p)$ satisfying (C1) and (C2),

$$\hat{m} \triangleq n^{\frac{1+\alpha(1-\hat{p})}{1+\alpha}}, \quad (2)$$

where $\alpha > 0$ is a tuning parameter. For fixed n , m is a monotone decreasing function of p , taking values in the interval $[n^{\frac{1}{1+\alpha}}, n]$. Thus, α governs the smallest acceptable resample size. The choice of α may be dictated (to be a fixed constant) by practical considerations (e.g. by setting the smallest acceptable resample size) or tuned using the data; we provide a double bootstrap procedure for choosing α below which we recommend for use in practice.

Another potentially important tuning parameter is $\tau_n(\mathbf{H}_{21})$. For a given patient history $\mathbf{h}_{21}$, the indicator $\mathcal{I}\{n[\mathbf{h}_{21}^T \hat{\psi}_{2,n}]^2 \leq \tau_n(\mathbf{h}_{21})\}$ can be viewed as the acceptance region for testing the null hypothesis $\mathbf{h}_{21}^T \psi_2 = 0$. Thus, a natural choice for $\tau_n(\mathbf{h}_{21})$ is $(\mathbf{h}_{21}^T \hat{\Sigma}_{21} \mathbf{h}_{21}) \cdot \chi_{1,1-\nu}^2$, where $\hat{\Sigma}_{21}$ is the plug-in estimator of $n \text{Cov}(\psi_{2,n}, \psi_{2,n})$ and $\chi_{1,1-\nu}^2$ is the $(1 - \nu) \times 100$ percentile of a χ^2 distribution with 1 degree of freedom. We use $\nu = 0.001$ in our simulation study. Robustness to the choice of ν is demonstrated in Web Appendix C.

Let $\hat{\mathbb{P}}_m^{(b)}$ denote the bootstrap empirical measure based on resamples of size m . For a given functional of the data, say $\hat{\rho}_n = f(\mathbb{P}_n)$, we write $\hat{\rho}_m^{(b)} = f(\hat{\mathbb{P}}_m^{(b)})$ to denote its m -out-of- n bootstrap analog. Let $\mathbf{c} \in \mathbb{R}^{\dim(\theta_1)}$ be arbitrary. To form a $(1-\eta) \times 100\%$ confidence interval for $\mathbf{c}^T \theta_1$, we first find $\hat{l}$ and $\hat{u}$, the $(\eta/2) \times 100$ and $(1 - \eta/2) \times 100$ percentiles of $\mathbf{c}^T \sqrt{\hat{m}}(\hat{\theta}_{1,\hat{m}}^{(b)} - \hat{\theta}_{1,n})$ respectively. The confidence set is then given by $(\mathbf{c}^T \hat{\theta}_{1,n} - \hat{u} / \sqrt{\hat{m}}, \mathbf{c}^T \hat{\theta}_{1,n} - \hat{l} / \sqrt{\hat{m}})$. We discuss the theoretical properties of this interval in Web Appendix B.

2.3.2 Double Bootstrap Algorithm for Choosing α —Here we describe a double bootstrap procedure (see Davison and Hinkley, 1997, for details) for choosing the tuning parameter α in a data-driven manner. In simulations, this procedure appears to reduce conservatism.

Suppose $\mathbf{c}^T \theta_1$ is the parameter of interest, and its estimate from the original data is $\mathbf{c}^T \hat{\theta}_{1,n}$. Consider a grid of candidate values for α ; we used $\{0.025, 0.05, 0.075, \dots, 1\}$ in our simulation study. The algorithm is as follows.

1. Draw B_1 n -out-of- n first-stage bootstrap samples from the data and calculate the bootstrap estimates $\mathbf{c}^T \hat{\theta}_{1,n}^{(b_1)}$, $b_1 = 1, \dots, B_1$. Fix α at the smallest value in the grid.
2. Compute the corresponding values of $m^{(b_1)}$ using equation (2), $b_1 = 1, \dots, B_1$.
3. Conditional on each first-stage bootstrap sample, draw B_2 $m^{(b_1)}$ -out-of- n second-stage (nested) bootstrap samples and calculate the double bootstrap versions of the estimate $\mathbf{c}^T \hat{\theta}_{1,n,\hat{m}^{(b_1)}}^{(b_1 b_2)}$, $b_1 = 1, \dots, B_1$, $b_2 = 1, \dots, B_2$.
4. For $b_1 = 1, \dots, B_1$, compute the $(\eta/2) \times 100$ and $(1 - \eta/2) \times 100$ percentiles of $\left\{ \mathbf{c}^T \sqrt{\hat{m}^{(b_1)}} \left(\hat{\theta}_{1,n,\hat{m}^{(b_1)}}^{(b_1 b_2)} - \hat{\theta}_{1,n}^{(b_1)} \right), b_2 = 1, \dots, B_2 \right\}$; say $\hat{l}_{DB}^{(b_1)}$ and $\hat{u}_{DB}^{(b_1)}$ respectively. Construct the *double centered percentile bootstrap* (Efron and Tibshirani, 1993) CI

from the b_1 -th first-stage bootstrap data as

$$(\mathbf{c}^T \hat{\theta}_{1,n}^{(b_1)} - \hat{u}_{\text{DB}}^{(b_1)} / \sqrt{\hat{m}^{(b_1)}}, \mathbf{c}^T \hat{\theta}_{1,n}^{(b_1)} - \hat{l}_{\text{DB}}^{(b_1)} / \sqrt{\hat{m}^{(b_1)}}), b_1=1, \dots, B_1, b_1=1, \dots, B_1.$$

5. Estimate the coverage rate of the double bootstrap CI from all the first-stage bootstrap data sets as

$$\frac{1}{B_1} \sum_{b_1=1}^{B_1} \mathcal{I} \left\{ \mathbf{c}^T \hat{\theta}_{1,n}^{(b_1)} - \hat{u}_{\text{DB}}^{(b_1)} / \sqrt{\hat{m}^{(b_1)}} \leq \mathbf{c}^T \hat{\theta}_{1,n} \leq \mathbf{c}^T \hat{\theta}_{1,n} - \hat{l}_{\text{DB}}^{(b_1)} / \sqrt{\hat{m}^{(b_1)}} \right\}.$$

6. If the above coverage rate is at or above the nominal level, then pick the current value of α as the final value. Otherwise, increment α to the next highest value in the grid.
7. Repeat steps 2 – 6, until the coverage rate of the double bootstrap CI, attains the nominal coverage rate, or the grid is exhausted.

In our empirical studies, with the foregoing grid values for α , the algorithm never terminated because of exhaustion of the grid. If this does occur, one should examine the observed values of p . If the values of p are concentrated close to zero, v may be increased; if not, the maximal value in the grid should be increased.

2.3.3 Extending the Bickel-Sakov Choice of m —Bickel and Sakov (2008) proposed an adaptive choice of m for constructing CIs for extrema. Although they did not consider the problem of DTRs, their method can be extended to construct CIs for stage-1 parameters in Q -learning; this extension is apt since the stage-1 parameters in Q -learning are functions of the maximum of stage-2 Q -functions. An algorithm to choose m in the context of Q -learning, adapted from their work is given below. As before, let $\mathbf{c}^T \theta_1$ be the parameter of interest.

1. Consider a sequence of m 's of the form: $m_j = \lceil q^j n \rceil$, for $j = 0, 1, 2, \dots$ and $0 < q < 1$, where $\lceil x \rceil$ denotes the smallest integer x ; q is a fixed tuning parameter.
2. For a given data set (with estimate $\mathbf{c}^T \hat{\theta}_{1,n}$), and for all j 's, define the following bootstrap empirical distribution function:

$$R_{m_j}^B(x, \mathbf{c}^T \hat{\theta}_{1,n}) = \frac{1}{B} \sum_{b=1}^B \mathcal{I} \left\{ \sqrt{m_j} \mathbf{c}^T (\hat{\theta}_{1,m_j}^{(b)} - \hat{\theta}_{1,n}) \leq x \right\},$$

where $\mathbf{c}^T \hat{\theta}_{1,m_j}^{(b)}$ is the b -th m_j -out-of- n bootstrap version of the estimate $\mathbf{c}^T \hat{\theta}_{1,n}$, $b = 1, \dots, B$.

3. Choose the data-adaptive choice of m , say m_{BS} , as the minimizer of the sup-norm of the successive differences between the bootstrap empirical distribution functions:

$$\hat{m}_{\text{BS}} = \arg \min_{m_j} \sup_x \left| R_{m_j}^B(x, \mathbf{c}^T \hat{\theta}_{1,n}) - R_{m_{j+1}}^B(x, \mathbf{c}^T \hat{\theta}_{1,n}) \right|.$$

Bickel and Sakov (2008) used $q = 0.75$ in their simulation study, and reported robustness to other values. We found that when their method is extended to Q -learning, the resultant confidence interval is sensitive to the choice of q and that larger values, say $q = 0.95$,

yielded better performance, while values smaller than 0.6 led to unstable results, possibly due to a small resample size; see Web Appendix D for a detailed sensitivity analysis.

3. Simulation Study

In this section, we use a simulation study to empirically evaluate the proposed m -out-of- n bootstrap methods. Nine generative models are used in these evaluations, each of them having two stages of treatment and two treatments at each stage. These models can be described in terms of the following components: (i) $O_j \in \{-1, 1\}$, $A_j \in \{-1, 1\}$ for $j = 1, 2$; (ii) $P(A_1 = 1) = P(A_1 = -1) = 0.5$, $P(A_2 = 1) = P(A_2 = -1) = 0.5$; (iii) $O_1 \sim \text{Bernoulli}(0.5)$, $O_2|O_1, A_1 \sim \text{Bernoulli}(\text{expit}(\delta_1 O_1 + \delta_2 A_1))$; (iv) $Y_1 \equiv 0$, $Y_2 = \gamma_1 + \gamma_2 O_1 + \gamma_3 A_1 + \gamma_4 O_1 A_1 + \gamma_5 A_2 + \gamma_6 O_2 A_2 + \gamma_7 A_1 A_2 + \varepsilon$, where $\varepsilon \sim N(0, 1)$ and $\text{expit}(x) = \exp(x)/(1 + \exp(x))$. This class is indexed by nine parameters $\gamma_1, \gamma_2, \dots, \gamma_7, \delta_1, \delta_2$. The analysis model uses the following patient history vectors: $\mathbf{H}_{20} = (1, O_1, A_1, O_1 A_1)^T$, $\mathbf{H}_{21} = (1, O_2, A_1)^T$, $\mathbf{H}_{10} = (1, O_1)^T$, $\mathbf{H}_{11} = (1, O_1)^T$. Our analysis models are given by $Q_2(\mathbf{H}_2, A_2; \beta_2, \psi_2) \triangleq \mathbf{H}_{20}^T \beta_2 + \mathbf{H}_{21}^T \psi_2 A_2$ and $Q_1(\mathbf{H}_1, A_1; \beta_1, \psi_1) \triangleq \mathbf{H}_{10}^T \beta_1 + \mathbf{H}_{11}^T \psi_1 A_1$.

This class of generative models, developed by Chakraborty et al. (2010), is useful as it allows us to influence the degree of nonregularity present in our example problems through the choice of γ 's and δ 's. Recall that in Q -learning, nonregularity occurs when more than one stage-2 treatment produces nearly the same optimal expected outcome for a set of patient histories that occur with positive probability. In the model class above, this occurs if the model generates histories for which $\gamma_5 A_2 + \gamma_6 O_2 A_2 + \gamma_7 A_1 A_2 \approx 0$, i.e., if it generates histories for which Q_2 depends weakly or not at all on A_2 . By manipulating the values of γ 's and δ 's, we can control: (i) the probability of generating a patient history such that $\gamma_5 A_2 + \gamma_6 O_2 A_2 + \gamma_7 A_1 A_2 = 0$, and (ii) the standardized effect size

$E[(\gamma_5 + \gamma_6 O_2 + \gamma_7 A_1) / \sqrt{\text{Var}(\gamma_5 + \gamma_6 O_2 + \gamma_7 A_1)}]$. Each of these quantities, denoted by p and ϕ , respectively, can be thought of as measures of nonregularity. Table 1 provides the parameter settings, classified as “nonregular”, “near-nonregular” and “regular.”

We compare the performances of the following methods for constructing CI for the stage-1 parameters of Q -learning: (i) the n -out-of- n bootstrap CI using the centered percentile bootstrap (CPB) method described in Section 2.3 (n -CPB); (ii) the adaptive confidence interval developed by Laber et al. (2012) using the resample size n (n -ACI); (iii) m -out-of- n bootstrap CI with the choice of m adapted from the work of Bickel and Sakov (2008) described in Section 2.3.3 with $q = 0.75$ and 0.95 ($m_{\text{BS},0.75}$ -CPB and $m_{\text{BS},0.95}$ -CPB); (iv) the proposed m -out-of- n bootstrap CIs with fixed $\alpha = 0.05$ and 0.1 , which correspond to minimum acceptable resample sizes of approximately 230 and 180 respectively ($m_{0.05}$ -CPB and $m_{0.1}$ -CPB); and (v) the proposed m -out-of- n bootstrap CIs with α chosen by double bootstrap ($m_{\hat{\alpha}}$ -CPB). We implemented the double bootstrap calculations with $B_1 = B_2 = 500$. We used $N = 1000$ simulated data sets and $B = 1000$ bootstrap replications, and fixed the sample size at $n = 300$. Additional simulation results for $n = 500$ and 1000 , as well as a sensitivity analysis with respect to the choice of α are provided in Web Appendix C.

3.1 Results

Average bootstrap resample sizes for all the methods are given in Table 2. The resample sizes of the adaptive m -out-of- n procedures are more variable than the Bickel-Sakov extensions. Tables 3 and 4 show the coverage rates and average widths for ψ_{10} , the main effect of treatment. As anticipated from the inconsistency of the n -out-of- n bootstrap, the n -CPB method exhibits under-coverage in most of the examples. The n -ACI, is a consistent bootstrap procedure that by design is conservative in the nonregular settings but delivers

nominal coverage rates as the settings become weakly nonregular or regular (Ex. 5, 6, A). The method adapted from the work of Bickel and Sakov (2008) is conservative across all the examples when $q = 0.75$ ($m_{BS,0.75}$ -CPB) but provides nominal coverage across all the examples when $q = 0.95$ ($m_{BS,0.95}$ -CPB). Our proposed m -out-of- n bootstrap method with fixed $\alpha = 0.05$ ($m_{\alpha,0.05}$ -CPB) provides nominal coverage across all the examples. However, when $\alpha = 0.1$ ($m_{\alpha,0.1}$ -CPB), its behavior becomes similar to that of n -ACI. The m -out-of- n bootstrap with data-driven α ($m_{\hat{\alpha}}$ -CPB) offers an extra layer of adaptiveness; by fine-tuning α it reduces the conservatism present in case of n -ACI and $m_{\alpha,0.1}$ -CPB and provides nominal coverage in all the examples. Reducing conservatism in nearly nonregular cases is of particular interest because these mimic the small signal-to-noise ratios common to sequential clinical trial data.

As expected, CIs constructed via the n -CPB method have the smallest average width; however, these are often associated with under-coverage. On the other hand, $m_{BS,0.75}$ -CPB method always gives the widest CIs. The widths of the CIs from the remaining five methods are comparable. Of these five methods, while the non-conservative methods ($m_{BS,0.95}$ -CPB, $m_{\alpha,0.05}$ -CPB, $m_{\hat{\alpha}}$ -CPB) generally outperform the conservative ones (n -ACI, $m_{\alpha,0.1}$ -CPB), there is no method that gives uniformly narrower CIs than the rest across all the examples.

The above findings establish that both the fixed α version and the data-driven α version of our proposed m -out-of- n bootstrap method, as well as a well-tuned version ($q = 0.95$) of the method due to Bickel and Sakov (2008), are at least as good as (and often better than) the n -ACI method in terms of finite-sample performance.

In terms of computational cost, the m -out-of- n bootstrap method with fixed α (either 0.05 or 0.1) is the cheapest. It is no more expensive than the usual n -out-of- n bootstrap; is approximately 20 times faster than either version ($q = 0.75$ or 0.95) of the method of Bickel and Sakov (2008); and is approximately 180 times faster than the ACI method. While the $m_{\hat{\alpha}}$ -CPB and the ACI are computationally comparable (each takes an average of 180 seconds to construct one CI on a 2.53 GHz Intel Core 2 Duo machine with 4 GB RAM), it is much easier to implement. Furthermore, the computational cost of the ACI regards the runtime of a heuristic solution to a nonconvex optimization problem (Laber et al., 2012); finding a global solution is presumably much more costly. The computational and conceptual simplicity might make the m -out-of- n bootstrap an attractive choice in practice.

4. Analysis of STAR*D Data

4.1 Study Details

Sequenced Treatment Alternatives to Relieve Depression (STAR*D) was a multisite, multilevel randomized controlled trial designed to assess the comparative effectiveness of different treatment regimes for patients with major depressive disorder (MDD). The study enrolled a total of 4041 patients, all of whom were treated with citalopram (CIT) at level 1.

Severity of depression was assessed using the clinician-rated and self-report versions of the *Quick Inventory of Depressive Symptomatology* (QIDS) scores (Rush et al., 2003). Success was based on a total clinician-rated QIDS-score of ≤ 5 ("remission") during treatment with CIT. Those without remission were eligible to receive one of up to seven treatment options available at level 2. Following Chakraborty and Moodie (2013), we classify the treatments into two categories: (i) Selective Serotonin Reuptake Inhibitor (SSRI), a common class of antidepressants (alone or in combination): sertraline (SER), CIT+bupropion (BUP), CIT+buspirone (BUS), or CIT + cognitive psychotherapy (CT); or (ii) non-SSRI: venlafaxine (VEN), BUP, or CT alone. Only the patients assigned CIT+CT or CT alone in level 2 were eligible, in case of a non-satisfactory response, to move to a supplementary level of

treatment (level 2A), to receive either VEN or BUP. Patients not responding satisfactorily at level 2 (and level 2A, if applicable) would continue to level 3 treatment. Treatment options at level 3 again can be classified into two categories: (i) SSRI: an augmentation of any SSRI-containing level 2 treatment with either lithium (Li) or thyroid hormone (THY); or (ii) non-SSRI: mirtazapine (MIRT) or nortriptyline (NTP), or an augmentation of any non-SSRI level 2 treatment with either Li or THY. Patients without a satisfactory response at level 3 continued to level 4 treatments, which included two options: tranylcypromine (TCP) or MIRT+VEN. Patients achieving remission (QIDS ≤ 5) at any level entered a follow-up phase. Treatment assignment at each level took place via randomization within a patient's preference category (preference to either switch or augment the treatment previously received, or no preference). For a complete description of the STAR*D study design, see Fava et al. (2003) and Rush et al. (2004). A schematic of the treatment assignment algorithm is given in Figure 1 of Web Appendix E.

4.2 Analysis

For the purpose of analysis, we consider level 2A as a part of level 2. A patient who received an SSRI at level 2 but a non-SSRI at level 2A is considered a recipient of SSRI in the combined level 2+2A. We treat levels 2 (including 2A, if applicable) and 3 as stages 1 and 2 respectively of the Q -learning framework, and do not consider level 4 data. Because of the trial design, at stage 2 we have the outcome data only for the non-remitters from stage 1; we define the overall primary outcome (Y) as the average $-QIDS$ score over the stage(s) a patient is present in the study, i.e.

$$Y = R_1 \cdot Y_1 + (1 - R_1) \cdot \left(\frac{Y_1 + Y_2}{2} \right),$$

where Y_1 and Y_2 are the $-QIDS$ scores measured at the end of stages 1 and 2 respectively (we take negatives to make higher values correspond to better outcomes), and $R_1 = 1$ if the subject achieves remission (QIDS ≤ 5) at the end of stage 1 and 0 otherwise. Following Pineau et al. (2007), we include three covariates in our analysis: (i) QIDS-score measured at the start of the level (QIDS.start), (ii) the slope of the QIDS-score over the previous level (QIDS.slope), and (iii) preference. While QIDS.start and QIDS.slope are continuous variables, preference is a binary variable, coded 1 for preference to switch previous treatment and -1 for preference to augment previous treatment or no preference. To conform to earlier notation, at the j th stage, for $j = 1, 2$ let: O_{1j} denote QIDS.start; O_{2j} denote QIDS.slope; O_{3j} denote preference; and A_j denote treatment, coded 1 for SSRI and -1 for non-SSRI. We use the following models for the Q -functions:

$$\begin{aligned} Q_2 &= \beta_{02} + \beta_{12}O_{12} + \beta_{22}O_{22} + \beta_{32}O_{32} + \beta_{42}A_1 + (\psi_{02} + \psi_{12}O_{12} + \psi_{22}O_{22})A_2, \\ Q_1 &= \beta_{01} + \beta_{11}O_{11} + \beta_{21}O_{21} + \beta_{31}O_{31} + (\psi_{01} + \psi_{11}O_{11} + \psi_{21}O_{21} + \psi_{31}O_{31})A_1. \end{aligned}$$

To avoid singularity, preference-by-treatment interaction is not included in the model for Q_2 ; similarly no A_1A_2 interaction is included. According to the above models, the optimal DTR is given by the following two decision rules: $d_2(\mathbf{H}_2) = \text{sign}(\psi_{02} + \psi_{12}O_{12} + \psi_{22}O_{22})$, and $d_1(\mathbf{H}_1) = \text{sign}(\psi_{01} + \psi_{11}O_{11} + \psi_{21}O_{21} + \psi_{31}O_{31})$. In our analysis, we use data on 1260 patients at stage 1 (level 2); a small number (19) of patients are omitted altogether due to gross item missingness in the covariates. Of the 1260 patients at stage 1, 792 were non-remitters (QIDS > 5) who should have moved to stage 2 (level 3); however, only 324 patients were present at stage 2 while the rest dropped out. To adjust for this dropout, we fit the model for Q_2 using *inverse probability weighting* where the probability of being present

at stage 2 is estimated by logistic regression using O_{11} , O_{21} , O_{31} , A_1 , $-Y_1$, O_{22} , $O_{11}A_1$, $O_{21}A_1$, and $O_{31}A_1$ as predictors. Another complexity arises in the computation of the pseudo-outcome, $\max_{a2} Q_2$. Note that for $(792 - 324) = 468$ non-remitters who were absent from stage 2, covariates O_{12} (QIDS.start at stage 2) and O_{32} (preference at stage 2) are missing, rendering the computation of the pseudo-outcome impossible for them. For these patients, we impute the value of O_{12} by the last observed QIDS score in the previous stage – a sensible strategy for a continuous, slowly changing variable like the QIDS score. On the other hand, we impute the missing values of the binary variable O_{32} (preference at stage 2) using k nearest neighbor (k -NN) classification, where k is chosen via leave-one-out cross-validation. Following these imputations, we implement Q -learning for this data; the estimates of the parameters of the Q -functions, along with their 95% bootstrap CIs are computed. While only the usual n -out-of- n bootstrap is used at stage 2, both the usual n -out-of- n bootstrap and the adaptive m -out-of- n bootstrap (with α chosen via double bootstrap) procedures are employed at stage 1, for ready comparison.

4.3 Results

Results of the above analysis are presented in Table 5. In our analysis, m is chosen as 1059 using double bootstrap. At both stages, the coefficient of QIDS.start (β_{12} and β_{11}) and the coefficient of preference (β_{32} and β_{31}) are significant at the 0.05 level. Additionally ψ_{31} , the coefficient of preference by treatment interaction at stage 1 are significant; this is interesting because it suggests that the decision rule at stage 1 should be individually tailored based on preference. The estimated optimal DTR can be explicitly described in terms of the ψ 's: $d_2(\mathbf{H}_2) = \text{sign}(-0.18 - 0.01O_{12} - 0.25O_{22})$, and $d_1(\mathbf{H}_1) = \text{sign}(-0.73 + 0.01O_{11} + 0.01O_{21} - 0.67O_{31})$. That is, the estimated optimal DTR suggests treating a patient at stage 2 with an SSRI if $(-0.18 - 0.01 \times \text{QIDS.start}_2 - 0.25 \times \text{QIDS.slope}_2) > 0$, and with a non-SSRI otherwise. Similarly, it suggests treating a patient at stage 1 with an SSRI if $(-0.73 + 0.01 \times \text{QIDS.start}_1 + 0.01 \times \text{QIDS.slope}_1 - 0.67 \times \text{preference}_1) > 0$, and with a non-SSRI otherwise.

The estimated difference in mean outcome at stage 2 corresponding to the two treatment options is given by $Q_2(\mathbf{H}_2, 1; \hat{\beta}_{2,n}, \hat{\psi}_{2,n}) - Q_2(\mathbf{H}_2, -1; \hat{\beta}_{2,n}, \hat{\psi}_{2,n}) = 2(-0.18 - 0.01 \times \text{QIDS.start}_2 - 0.25 \times \text{QIDS.slope}_2)$. Likewise, the estimated difference in mean pseudo-outcome at stage 1 corresponding to the two treatment options is given by $Q_1(\mathbf{H}_1, 1; \hat{\beta}_{1,n}, \hat{\psi}_{1,n}) - Q_1(\mathbf{H}_1, -1; \hat{\beta}_{1,n}, \hat{\psi}_{1,n}) = 2(-0.73 + 0.01 \times \text{QIDS.start}_1 + 0.01 \times \text{QIDS.slope}_1 - 0.67 \times \text{preference}_1)$. For any fixed values of QIDS.start, QIDS.slope, and preference, one can construct point-wise CIs for the above difference in mean outcome (or, pseudo-outcome) based on the CIs for the individual ψ 's, thus leading to a confidence band around the entire function. The mean difference function and its 95% confidence band over the observed range of QIDS.start and QIDS.slope are plotted for stage 1 (separately for preference = “switch” and preference = “augment or no preference”) and for stage 2 (patients with all preference categories combined), and are presented in Figure 2 of Web Appendix E. It turns out that the confidence bands in all three panels contain zero, thereby suggesting insufficient evidence in the data to recommend one uniquely best treatment.

5. Discussion

We proposed an adaptive m -out-of- n bootstrap scheme for constructing confidence intervals for coefficients indexing an optimal dynamic treatment regime. The confidence intervals were shown to deliver asymptotically valid coverage under fixed alternatives and displayed competitive finite-sample performance in simulations. The intervals have the advantage of being easy to understand, simple to program, and fast to execute. We hope that such intervals will prove more palatable to practitioners who may find procedures like the ACI

(Laber et al., 2012) and PQ -learning (Song et al., 2012) too complex or too burdensome computationally.

It is worth noting that while we used an adaptive m -out-of- n resampling scheme, we could have instead used subsampling without replacement. In particular, a subsample size of $m \approx m/2$ would enjoy similar asymptotic theory to the adaptive m -out-of- n bootstrap and hence provide consistent confidence sets (see, for example, Politis et al., 1999). Furthermore, with this choice of subsample size, under a completely regular generative model, $m \approx n/2$ which closely mimics the bootstrap (Friedman and Hall, 2007). One possible advantage of the adaptive m -out-of- n scheme over an adaptive subsampling scheme is that in a regular setting, the m -out-of- n procedure reduces to the usual n -out-of- n bootstrap which may be more familiar to both researchers and practitioners.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Bibhas Chakraborty acknowledges support from the NIH grant R01 NS072127-01A1. Eric Laber acknowledges support from NIH grant P01 CA142538. Data used in the preparation of this article were obtained from the limited access datasets distributed from the NIMH. The STAR*D study was supported by NIMH Contract # N01MH90003 to the University of Texas Southwestern Medical Center. The [ClinicalTrials.gov](https://clinicaltrials.gov/ct2/show/study/NCT00021528) identifier is NCT00021528. We are grateful to the editors and the reviewers for their insightful comments which have led to important improvements in this paper.

This manuscript reflects the views of the authors and may not reflect the opinions or views of the STAR*D Study Investigators or the NIH.

References

- Bellman, R. Dynamic programming. Princeton: Princeton University Press; 1957.
- Bickel P, Freedman D. Some asymptotic theory for the bootstrap. *The Annals of Statistics*. 1981; 9:1196–1217.
- Bickel P, Gotze F, van Zwet W. Resampling fewer than n observations: gains, losses and remedies for losses. *Statistica Sinica*. 1997; 7:1–31.
- Bickel P, Sakov A. On the choice of m in the m out of n bootstrap and confidence bounds for extrema. *Statistica Sinica*. 2008; 18:967–985.
- Bretagnolle, J. Annales de l'IHP Probabilités et statistiques. Vol. volume 19. Elsevier; 1983. Lois limites du bootstrap de certaines fonctionnelles. In; p. 281-296.
- Chakraborty B, Moodie E. Estimating optimal dynamic treatment regimes with shared decision rules across stages: An extension of q -learning. 2013 Submitted.
- Chakraborty B, Murphy S, Strecher V. Inference for non-regular parameters in optimal dynamic treatment regimes. *Statistical Methods in Medical Research*. 2010; 19:317–343. [PubMed: 19608604]
- Cheung K, Lee S, Young G. Iterating the m out of n bootstrap in nonregular smooth function models. *Statistica Sinica*. 2005; 15:945–967.
- Davison, A.; Hinkley, D. Bootstrap methods and their application. Cambridge, UK: Cambridge University Press; 1997.
- Dawson R, Lavori P. Placebo-free designs for evaluating new mental health treatments: the use of adaptive treatment strategies. *Statistics in Medicine*. 2004; 23:3249–3262. [PubMed: 15490427]
- Dumbgen L. On nondifferentiable functions and the bootstrap. *Probability Theory and Related Fields*. 1993; 95:125–140.
- Efron B. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*. 1979; 7:1–26.

- Efron, B.; Tibshirani, R. An Introduction to the Bootstrap. Vol. volume 57. Chapman & Hall/CRC; 1993.
- Fava M, Rush A, Trivedi M, Nierenberg A, Thase M, Sackeim H, Quitkin F, Wisniewski S, Lavori P, Rosenbaum J, Kupfer D. Background and rationale for the sequenced treatment alternatives to relieve depression (star*d) study. *Psychiatric Clinics of North America*. 2003; 26:457–494. [PubMed: 12778843]
- Friedman J, Hall P. On bagging and nonlinear estimation. *Journal of statistical planning and inference*. 2007; 137:669–683.
- Hall P, Horowitz J, Jing B. On blocking rules for the bootstrap with dependent data. *Biometrika*. 1995; 82:561–574.
- Hirano K, Porter J. Impossibility results for nondifferentiable functionals. *Econometrica*. 2012 forthcoming.
- Huang J, Sen P, Shao J. Bootstrapping a sample quantile when the density has a jump. *Statistica Sinica*. 1996; 6:299–309.
- Laber E, Murphy S. Adaptive confidence intervals for the test error in classification. *Journal of the American Statistical Association*. 2011; 106:904–913. [PubMed: 22053123]
- Laber E, Qian M, Lizotte D, Murphy S. Statistical inference in dynamic treatment regimes. 2012 Submitted.
- Lavori P, Dawson R. A design for testing clinical strategies: biased adaptive within-subject randomization. *Journal of the Royal Statistical Society, Series A*. 2000; 163:29–38.
- Lavori P, Dawson R. Dynamic treatment regimes: practical design considerations. *Clinical Trials*. 2004; 1:9–20. [PubMed: 16281458]
- Lee S. On a class of m out of n bootstrap confidence intervals. *Journal of the Royal Statistical Society, Series B*. 1999; 61:901–911.
- Moodie E, Richardson T. Estimating optimal dynamic regimes: correcting bias under the null. *Scandinavian Journal of Statistics*. 2010; 37:126–146.
- Murphy S. Optimal dynamic treatment regimes (with discussions). *Journal of the Royal Statistical Society, Series B*. 2003; 65:331–366.
- Murphy S. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. 2005a; 24:1455–1481. [PubMed: 15586395]
- Murphy S. A generalization error for q-learning. *Journal of Machine Learning Research*. 2005b; 6:1073–1097. [PubMed: 16763665]
- Pineau J, Bellare M, Rush A, Ghizaru A, Murphy S. Constructing evidence-based treatment strategies using methods from computer science. *Drug and Alcohol Dependence*. 2007; 88:S52–S60. [PubMed: 17320311]
- Politis, D.; Romano, J.; Wolf, M. Subsampling. Springer Verlag; 1999.
- Pötscher B. Confidence sets based on sparse estimators are necessarily large. Arxiv preprint arXiv: 0711.1036. 2007
- Pötscher, B.; Schneider, U. Mpra paper. Germany: University Library of Munich; 2008. Confidence sets based on penalized maximum likelihood estimators.
- Robins, J. Optimal structural nested models for optimal sequential decisions. In: Lin, D.; Heagerty, P., editors. *Proceedings of the Second Seattle Symposium on Biostatistics*. New York: Springer; 2004. p. 189–326.
- Rush A, Fava M, Wisniewski S, Lavori P, Trivedi M, Sackeim H, Thase M, Nierenberg A, Quitkin F, Kashner T, Kupfer D, Rosenbaum J, Alpert J, Stewart J, McGrath P, Biggs M, Shores-Wilson K, Lebowitz B, Ritz L, Niederehe G. Sequenced treatment alternatives to relieve depression (star*d): rationale and design. *Controlled Clinical Trials*. 2004; 25:119–142. [PubMed: 15061154]
- Rush A, Trivedi M, Ibrahim H, Carmody T, Arnow B, Klein D, Markowitz J, Ninan P, Kornstein S, Manber R, Thase M, Kocsis J, Keller M. The 16-item quick inventory of depressive symptomatology (qids), clinician rating (qids-c), and self-report (qids-sr): a psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*. 2003; 54:573–583. [PubMed: 12946886]

- Schulte P, Tsiatis A, Laber E, Davidian M. Q- and a-learning methods for estimating optimal dynamic treatment regimes. Arxiv preprint arXiv:1202.4177. 2012
- Shao J. Bootstrap sample size in nonregular cases. Proceedings of the American Mathematical Society. 1994; 122:1251–1262.
- Shao J, Wu C. A general theory for jackknife variance estimation. The Annals of Statistics. 1989:1176–1197.
- Shortreed S, Laber E, Lizotte D, Stroup T, Pineau J, Murphy S. Informing sequential clinical decision-making through reinforcement learning: an empirical study. Machine Learning. 2011; 84:109–136. [PubMed: 21799585]
- Song R, Wang W, Zeng D, Kosorok M. Penalized q-learning for dynamic treatment regimes. 2012 Submitted.
- Sutton, R.; Barto, A. Reinforcement learning: An introduction. Cambridge: MIT Press; 1998.
- Swanepoel J. A note on proving that the (modified) bootstrap works. Communications in Statistics-Theory and Methods. 1986; 15:3193–3203.
- Watkins, C. Ph.D. Thesis. Cambridge University; 1989. Learning from delayed rewards.
- Zhao Y, Kosorok M, Zeng D. Reinforcement learning design for cancer clinical trials. Statistics in Medicine. 2009; 28:3294–3315. [PubMed: 19750510]

Table 1

Parameters indexing the example models. Examples 1–6 were developed by Chakraborty et al. (2010), and examples A–C were developed by Laber et al. (2012).

Example	γ	δ	Type	Regularity	Measures
1	$(0, 0, 0, 0, 0, 0, 0)^T$	$(0.5, 0.5)^T$	nontregular	$p = 1$	$\phi = 0/0$
2	$(0, 0, 0, 0, 0.01, 0, 0)^T$	$(0.5, 0.5)^T$	near-nontregular	$p = 0$	$\phi = \infty$
3	$(0, 0, -0.5, 0, 0.5, 0, 0.5)^T$	$(0.5, 0.5)^T$	nontregular	$p = 1/2$	$\phi = 1.0$
4	$(0, 0, -0.5, 0, 0.5, 0, 0.49)^T$	$(0.5, 0.5)^T$	near-nontregular	$p = 0$	$\phi = 1.02$
5	$(0, 0, -0.5, 0, 1.0, 0.5, 0.5)^T$	$(1.0, 0.0)^T$	nontregular	$p = 1/4$	$\phi = 1.41$
6	$(0, 0, -0.5, 0, 0.25, 0.5, 0.5)^T$	$(0.1, 0.1)^T$	regular	$p = 0$	$\phi = 0.35$
A	$(0, 0, -0.25, 0, 0.75, 0.5, 0.5)^T$	$(0.1, 0.1)^T$	regular	$p = 0$	$\phi = 1.035$
B	$(0, 0, 0, 0, 0.25, 0, 0.25)^T$	$(0, 0)^T$	nontregular	$p = 1/2$	$\phi = 1.00$
C	$(0, 0, 0, 0, 0.25, 0, 0.24)^T$	$(0, 0)^T$	near-nontregular	$p = 0$	$\phi = 1.03$

Table 2

Average bootstrap resample size over 1000 simulated data sets. The rows represent different methods of constructing CIs: (i) the n-out-of-n centered percentile bootstrap (n-CPB); (ii) the adaptive confidence interval developed by Laber et al. (2012) using the resample size n (n-ACI); (iii – iv) m-out-of-n bootstrap with the choice of m adapted from the work of Bickel and Sakov (2008) with $q = 0.75$ and 0.95 ($\hat{m}_{BS,0.75}$ -CPB and $\hat{m}_{BS,0.95}$ -CPB); (v – vi) the proposed m-out-of-n bootstrap with fixed $\alpha = 0.05$ and 0.1 ($m_{0.05}$ -CPB and $m_{0.1}$ -CPB); and (vii) the proposed m-out-of-n bootstrap with a data-driven choice of α chosen by double bootstrap ($\hat{m}_\alpha$ -CPB). Average values of the tuning parameter α are given within parenthesis for the last three methods. Examples (along the columns) are designated NR = nonregular, NNR = near-nonregular, R = regular.

$n=300$	Ex. 1 NR	Ex. 2 NNR	Ex. 3 NR	Ex. 4 NNR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NNR
n -CPB	300	300	300	300	300	300	300	300	300
n -ACI	300	300	300	300	300	300	300	300	300
$\hat{m}_{BS,0.75}$ -CPB	175.47	174.46	182.87	177.32	178.05	173.19	180.41	169.30	169.91
$\hat{m}_{BS,0.95}$ -CPB	260.09	260.75	259.44	259.18	259.90	260.56	259.55	260.92	260.94
$m_{0.05}$ -CPB	229.12 (0.05)	229.09 (0.05)	262.49 (0.05)	262.49 (0.05)	280.92 (0.05)	271.58 (0.05)	284.29 (0.05)	260.55 (0.05)	260.24 (0.05)
$m_{0.1}$ -CPB	179.18 (0.1)	179.14 (0.1)	232.14 (0.1)	232.15 (0.1)	264.13 (0.1)	247.93 (0.1)	270.54 (0.1)	228.97 (0.1)	228.45 (0.1)
$\hat{m}_\alpha$ -CPB	208.61 (0.081)	208.80 (0.081)	262.72 (0.055)	260.12 (0.059)	283.47 (0.046)	259.02 (0.080)	287.49 (0.040)	254.21 (0.065)	253.67 (0.066)

Table 3

Monte Carlo estimates of coverage rates of confidence intervals for the main effect of treatment (ψ_{10}) at the 95% nominal level. The rows represent different methods of constructing CIs: (i) the n-out-of-n centered percentile bootstrap (n-CPB); (ii) the adaptive confidence interval developed by Laber et al. (2012) using the resample size n (n-ACI); (iii) – iv) m-out-of-n bootstrap with the choice of m adapted from the work of Bickel and Sakov (2008) with $q = 0.75$ and 0.95 ($m_{BS,0.75}$ -CPB and $m_{BS,0.95}$ -CPB); (v – vi) the proposed m-out-of-n bootstrap with fixed $\alpha = 0.05$ and 0.1 ($m_{0.05}$ -CPB and $m_{0.1}$ -CPB); and (vii) the proposed m-out-of-n bootstrap with a data-driven choice of α chosen by double bootstrap ($m_{\hat{\alpha}}$ -CPB).

$n=300$	Ex. 1 NR	Ex. 2 NR	Ex. 3 NR	Ex. 4 NR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NR
n -CPB	0.936	0.932 *	0.928 *	0.921 *	0.933 *	0.931 *	0.944	0.925 *	0.922 *
n -ACI	0.994	0.994	0.975	0.976	0.962	0.957	0.950	0.977	0.976
$m_{BS,0.75}$ -CPB	0.979	0.98	0.98	0.977	0.969	0.976	0.975	0.982	0.977
$m_{BS,0.95}$ -CPB	0.954	0.954	0.954	0.955	0.954	0.943	0.951	0.948	0.952
$m_{0.05}$ -CPB	0.960	0.963	0.944	0.945	0.936	0.941	0.946	0.951	0.952
$m_{0.1}$ -CPB	0.984	0.982	0.956	0.955	0.943	0.949	0.953	0.971	0.970
$m_{\hat{\alpha}}$ -CPB	0.964	0.964	0.953	0.950	0.939	0.947	0.944	0.955	0.960

Coverage rates significantly different from 0.95 at the 0.05 level are in bold. Of these significantly different-from-nominal coverage rates, the ones that are significantly below 0.95 (denoting more serious problem due to non-regularity) are marked with *.

Examples (along the columns) are designated NR = nonregular, NNR = near-nonregular, R = regular.

Table 4

Monte Carlo estimates of the mean width of confidence intervals for the main effect of treatment (ψ_{10}) at the 95% nominal level. The rows represent different methods of constructing CIs: (i) the n-out-of-n centered percentile bootstrap (n-CPB); (ii) the adaptive confidence interval developed by Laber et al. (2012) using the resample size n (n-ACI); (iii) – iv) m-out-of-n bootstrap with the choice of m adapted from the work of Bickel and Sakov (2008) with $q = 0.75$ and 0.95 ($m_{BS,0.75}$ -CPB and $m_{BS,0.95}$ -CPB); (v – vi) the proposed m-out-of-n bootstrap with fixed $\alpha = 0.05$ and 0.1 ($m_{0.05}$ -CPB and $m_{0.1}$ -CPB); and (vii) the proposed m-out-of-n bootstrap with a data-driven choice of α chosen by double bootstrap ($m_{\hat{\alpha}}$ -CPB).

$n=300$	Ex. 1 NR	Ex. 2 NR	Ex. 3 NR	Ex. 4 NR	Ex. 5 NR	Ex. 6 R	Ex. A R	Ex. B NR	Ex. C NR
n -CPB	0.269	0.269*	0.300*	0.300*	0.320*	0.309*	0.314	0.299*	0.299*
n -ACI	0.354	0.354	0.342	0.342	0.341	0.327	0.327	0.342	0.342
$m_{BS,0.75}$ -CPB	0.402	0.400	0.431	0.441	0.473	0.46	0.464	0.455	0.455
$m_{BS,0.95}$ -CPB	0.289	0.288	0.324	0.324	0.345	0.334	0.339	0.322	0.322
$m_{0.05}$ -CPB	0.306	0.306	0.320	0.320	0.331	0.325	0.323	0.321	0.321
$m_{0.1}$ -CPB	0.346	0.347	0.341	0.341	0.340	0.341	0.332	0.342	0.343
$m_{\hat{\alpha}}$ -CPB	0.331	0.331	0.321	0.323	0.330	0.336	0.322	0.328	0.328

Widths with corresponding coverage rates significantly different from the nominal rate are in bold, and those with corresponding coverage rates significantly below the nominal rate are marked with *. Examples are designated NR = nonregular, NNR = near-nonregular, R = regular.

Table 5

Regression coefficients and their 95% centered percentile bootstrap CIs (both the usual n-out-of-n and the novel m-out-of-n) in the analysis of STAR*D data (significant coefficients are in bold).

Parameter	Variable	Estimate	95% CI (<i>n</i> -out-of- <i>n</i>)	95% CI (<i>m</i> -out-of- <i>n</i>)
Stage 2 (<i>n</i> = 324)				
β_{02}	Intercept ₂	-1.66	(-3.70, 0.43)	–
β_{12}	QIDS.start ₂	-0.72	(-0.87, -0.56)	–
β_{22}	QIDS.slope ₂	0.79	(-0.32, 1.99)	–
β_{32}	Preference ₂	0.74	(0.05, 1.50)	–
β_{42}	Treatment ₁	0.26	(-0.38, 0.89)	–
ψ_{02}	Treatment ₂	-0.18	(-2.15, 2.00)	–
ψ_{12}	Treatment ₂ × QIDS.start ₂	-0.01	(-0.18, 0.13)	–
ψ_{22}	Treatment ₂ × QIDS.slope ₂	-0.25	(-1.33, 0.94)	–
Stage 1 (<i>n</i> = 1260; <i>m</i> ≐ 1059)				
β_{01}	Intercept ₁	-0.47	(-1.64, 0.71)	(-1.82, 0.97)
β_{11}	QIDS.start ₁	-0.55	(-0.63, -0.48)	(-0.65, -0.46)
β_{21}	QIDS.slope ₁	0.12	(-0.36, 0.52)	(-0.41, 0.57)
β_{31}	Preference ₁	0.88	(0.40, 1.40)	(0.35, 1.46)
ψ_{01}	Treatment ₁	-0.73	(-1.84, 0.43)	(-1.91, 0.48)
ψ_{11}	Treatment ₁ × QIDS.start ₁	0.01	(-0.06, 0.09)	(-0.07, 0.09)
ψ_{21}	Treatment ₁ × QIDS.slope ₁	0.01	(-0.44, 0.46)	(-0.47, 0.49)
ψ_{31}	Treatment ₁ × Preference ₁	-0.67	(-1.17, -0.18)	(-1.29, -0.16)