University of California, Berkeley
U.C. Berkeley Division of Biostatistics Working Paper Series

Year 2014 *Paper* 330

Online Targeted Learning

Mark J. van der Laan* Samuel D. Lendle†

*Division of Biostatistics, University of California, Berkeley, laan@berkeley.edu
†Division of Biostatistics, University of California, Berkeley, lendle@stat.berkeley.edu

Online Targeted Learning

Mark J. van der Laan and Samuel D. Lendle

Abstract

We consider the case that the data comes in sequentially and can be viewed as sample of independent and identically distributed observations from a fixed data generating distribution. The goal is to estimate a particular path wise target parameter of this data generating distribution that is known to be an element of a particular semi-parametric statistical model. We want our estimator to be asymptotically efficient, but we also want that our estimator can be calculated by updating the current estimator based on the new block of data without having to revisit the past data, so that it is computationally much faster to compute than recomputing a fixed estimator each time new data comes in. We refer to such an estimator as an online estimator. These online estimators can also be applied on a large fixed data base by dividing the data set in many subsets and enforcing an ordering of these subsets. The current literature provides such online estimators for parametric models, where the online estimators are based on variations of the stochastic gradient descent algorithm.

For that purpose we propose a new online one-step estimator, which is proven to be asymptotically efficient under regularity conditions. This estimator takes as input online estimators of the relevant part of the data generating distribution and the nuisance parameter that are required for efficient estimation of the target parameter. These estimators could be an online stochastic gradient descent estimator based on large parametric models as developed in the current literature, but we also propose other online data adaptive estimators that do not rely on the specification of a particular parametric model.

We also present a targeted version of this online one-step estimator that presumably minimizes the one-step correction and thereby might be more robust in finite samples. These online one-step estimators are not a substitution estimator and might therefore be unstable for finite samples if the target parameter is borderline

identifiable.

Therefore we also develop an online targeted minimum loss-based estimator, which updates the initial estimator of the relevant part of the data generating distribution by updating the current initial estimator with the new block of data, and estimates the target parameter with the corresponding plug-in estimator. The online substitution estimator is also proven to be asymptotically efficient under the same regularity conditions required for asymptotic normality of the online one-step estimator.

The online one-step estimator, targeted online one-step estimator, and online TMLE is demonstrated for estimation of a causal effect of a binary treatment on an outcome based on a dynamic data base that gets regularly updated, a common scenario for the analysis of electronic medical record data bases.

Finally, we extend these online estimators to a group sequential adaptive design in which certain components of the data generating experiment are continuously fine-tuned based on past data, and the new data generating distribution is then used to generate the next block of data.

1 Introduction

This paper concerns online semi-parametric efficient estimation of a target parameter based on an online data base that is regularly updated with additional data for a new sub-sample of units. The proposed procedure needs to be able to provide the most up to date estimator based on all currently available data continuously in time, and thus not be significantly slowed down by the computational or memory burden, but we want to achieve this without giving up on statistical performance as measured by asymptotic efficiency, coverage of confidence intervals, and good finite sample behavior.

Semiparametric efficient estimators of a pathwise differentiable target parameter have been developed in the literature, using general approaches such as one-step estimation (Bickel et al. (1997)), estimating equation methodology (Robins and Rotnitzky (1992); van der Laan and Robins (2003)), and targeted minimum loss-based plug-in estimation (van der Laan and Rubin, 2006; Rose and van der Laan, 2011). However, these estimators were not restricted to be fast enough to be applied to a dynamic data base continuously in time. In addition, nowadays, data bases can be truly massive, possibly containing around 10^9 units, so that even the single application of a complex asymptotically efficient estimator can be computationally intractable.

In the machine learning literature, large scale online estimation problems are sometimes addressed with stochastic gradient descent (SGD) algorithms that approximate the computation of a minimum of an empirical risk over a finite dimensional parameter space. These optimization routines operate on a single observation or relatively small "mini-batches" of observations at a time. For example, SGD in classification or prediction problems aim to minimize some (possibly regularized) empirical risk as in logistic regression or a support vector machine (Bottou, 2010) over a finite dimensional parameter (possibly very high dimensional). Though a stochastic gradient descent algorithm takes far more steps to converge than other optimization routines which operate on the full data set at each step, statistically, SGD and some variants can perform well with only a single pass through a data set(Murata, 1998; Xu, 2011). Thus, SGD type methods can be useful as online estimators for finite dimensional parameters that can be expressed as an optimum of an empirical risk. In this manuscript we aim to develop online semiparametric efficient estimators for any pathwise differentiable target parameter in general semiparametric models which are thereby also an estimator for massive data sets by applying them to an ordered partitioning of the data set.

Firstly, we will propose a general template for online one-step estimation and online targeted minimum loss-based estimation that satisfies these stringent computational constraints while it is still asymptotically efficient and expected to be highly competitive with an unconstrained (intractable) one-step estimator or TMLE based on finite samples. Subsequently, we aim to extend this framework to handle an online data base in which certain components of the data generating distribution are changed over time.

1.1 Organization of article

The organization of this article is as follows. In the next subsection we provide some review of relevant literature on stochastic gradient descent algorithms for parametric working models. In Section 2 we define the general online estimation problem addressed in this article in terms of a (general) statistical model and target parameter, a formal definition of an online estimator, the efficient influence curve and asymptotic efficiency of an estimator. In Section 3 we define the online one-step estimator and establish its asymptotic consistency, normality, and efficiency, under regularity conditions, where we rely on a martingale central limit theorem for discrete martingales. In Section 4 we present targeted online one-step estimators that are asymptotically equivalent with the average across all the batches of the most up to date (at that batch) substitution estimator under a condition that the targeting succeeds in making the online empirical mean of the efficient influence curve asymptotically negligible. We propose an algorithm that aims to succeed in the latter, but no formal proof is provided.

In Section 5 we define an online targeted minimum loss-based estimator and establish its asymptotic normality and efficiency under the same regularity conditions as needed for the one-step estimator. In Section 6 we define online data adaptive initial estimators (including, online super-learner) that can be inputted in our online one-step and online TMLE of the target parameter. In Section 7 we show a particular application of our online one-step estimator to estimate the finite dimensional parameter of a working parametric model, thereby showing it provides an asymptotically equivalent (and efficient) alternative to the second order online SGD algorithms in the current literature. In Section 8 we work out our online one-step estimator to estimate the additive causal effect of a binary treatment on an outcome of interest, controlling for a set of baseline confounders. In Section 9 we do the same for the online TMLE and online targeted one-step estimator. In Section 10 we define the more general group sequential data adaptive design setting for this general online estimation problem, where the controlled censoring or treatment component of the data generating distribution can be fine tuned based on past data, and thus change over time. In Section 10, we also generalize the online one-step estimator and its theory to the adaptive group sequential design setting, and clarify that the same generalization applies to the other proposed online estimators. We conclude with a discussion in Section 11. Some proofs are presented in the Appendix.

1.2 Some review of literature on stochastic gradient descent optimization for parametric working models

Here we describe the stochastic gradient descent optimizer and review some results from the literature. Suppose one observes n independent and identically distributed $O_1, \ldots, O_n$ with common probability distribution P_0, and define the parameter of interest as

$$\theta_0 = \underset{\theta}{\operatorname{argmin}}\ P_0 L(\theta)$$

for $\theta \in \mathbb{R}^d$, where $O \to L(\theta)(O)$ is a loss function of $O \sim P_0$, and we used the notation $P_0 f = \int f(o) dP_0(o)$ for the expectation operator. For example, if

$$L(\theta)(O) = \frac{1}{2}(Y - \theta' X)^2,$$

where $O = (X, Y)$, $X \in \mathbb{R}^d$, and $Y \in \mathbb{R}$, this is the ordinary least squares regression problem. One can also include a regularization term in the loss function, and other examples include generalized linear models, and support vector machines.

For a data set with empirical distribution P_n, call the true optimum of the empirical mean of the loss function, also known as the empirical risk, $\hat{\theta}_n$. That is,

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\ P_n L(\theta) = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} L(\theta)(O_i).$$

When $L(\theta) = -\log p_\theta$ for some parametric model $\{p_\theta : \theta \in \Theta\}$, $\hat{\theta}_n$ is the maximum likelihood estimator. Let $V_\theta = P_0 \frac{d^2}{d\theta^2} L(\theta)$. Under mild regularity conditions (see e.g., (van der Vaart, 1998)), we have $\hat{\theta}_n$ is asymptotically normally distributed with asymptotic variance

$$V_{\theta_0}^{-1} P_0 \left[\frac{d}{d\theta_0} L(\theta_0)^\top \frac{d}{d\theta_0} L(\theta_0) \right] V_{\theta_0}^{-1}.$$

Stochastic gradient descent is an iterative optimization routine which takes a small step in the direction of a single randomly selected observation from the data set. In practice, the data set is usually shuffled or assumed to be in random order and processed sequentially. Let

$$\theta_{t+1} = \theta_t + \gamma_t \Gamma_t \frac{d}{d\theta_t} L(\theta_t)(O_t) \tag{1}$$

where γ_t is a scalar step size or learning rate, Γ_t is a $d \times d$ matrix, and O_t is the observation used at the t-th step (Bottou, 2010). After some number of steps, we hope that θ_t is sufficiently close to the true optimum $\hat{\theta}_n$ of the empirical risk. In particular, we hope that n steps is enough so that the SGD estimate θ_n after a single pass through the data set is a reasonable estimate of θ_0.

In the simplest version of SGD Γ_t is some constant times the identity matrix. Other variants replace Γ_t with an appropriate diagonal matrix (e.g., diagonal elements of $V_{\theta_t}^{-1}$) as in Adagrad (Duchi et al., 2011) and Adadelta (Zeiler, 2012), which are methods that tend to work well in practice. Murata (1998) shows that the mean and variance of θ_t depend on the learning rate γ_t and the eigenvalues of the matrix $\Gamma_t V_{\hat{\theta}_n}$. Second order SGD takes the curvature of the loss function into account, using some Γ_t that approximates $V_{\hat{\theta}_n}^{-1}$. Murata (1998) shows that when $\Gamma_t = V_{\hat{\theta}_n}^{-1}$ and γ_t is asymptotically $1/t$, θ_n, the second order SGD estimate after a single pass through the data set, is asymptotically equivalent with the true empirical optimum $\hat{\theta}_n$. That is, asymptotically, the variance of second order SGD divided by the variance of $\hat{\theta}_n$ converges to 1 as $n \to \infty$. Murata (1998) shows that, if Γ_t is constant and some weak conditions hold, then θ_n has bias of $O(1/n^{\lambda_d})$, where λ_d is the smallest eigenvalue of $\Gamma_n V_{\hat{\theta}_n}$, and the variance is $O(1/n)$ if $\lambda_d > 1/2$.

Though optimal, due to the high dimension of d, second order SGD is rarely used in practice because it is often to expensive to compute and store (an estimate of) $V_{\theta_n}^{-1}$. Averaged stochastic gradient descent (ASGD) is another different but related method to SGD which is very simple to implement. The ASGD estimate at step t is simply

$$\bar{\theta}_t = \frac{1}{t} \sum_{i=1}^{t} \theta_i$$

where θ_i is the SGD estimate at step i as in (1), Γ_t is the identity matrix times a constant, and γ_t now goes to 0 slower than $1/t$. (Polyak and Juditsky, 1992; Xu, 2011) show that in a single pass through the data set, $\bar{\theta}_n$ is also asymptotically optimal and thus equivalent with $\hat{\theta}_n$. Xu (2011) note that ASGD is not frequently used in practice possibly due to required tuning and the possibly huge number of observations required to reach the asymptotic performance, but it is shown in simulations that with some careful tuning, ASGD can perform very well.

There are many other variants to stochastic gradient descent type optimization routines. For more information and some insightful notes on implementation details, see (Bottou, 2012) and references therein.

2 Formulation of the online estimation problem

Let $O_1, \ldots, O_n$ be a set of n i.i.d. observations with probability distribution $P_0 \in \mathcal{M}$ where $\mathcal{M}$ denotes the statistical model (i.e., a collection of probability distribution that is known to contain the true one). Let $0 = n_0 < n_1 < n_2 < \ldots < n_K = n$, and we refer to this number of batches K also as $K(n)$. Let m be an upper bound for the sample size $n_j - n_{j-1}$ of the j-th batch so that $\max_{j=1,\ldots,K} n_j - n_{j-1} \leq m$. For sake of presentation and notational convenience, we will actually assume $n_j - n_{j-1} = m$ is constant in j so that $n = Km$, but all estimators have straightforward extensions to the case that $m_j = n_j - n_{j-1}$ can vary over j and is bounded from above by some m.

Here $n = n_K$ represents the current sample size, while n_j represents the sample size reached at the j–th stage, where each stage j adds a next group of $n_j - n_{j-1}$ observations O_i with $i = n_{j-1} + 1, \ldots, n_j$. We are not assuming that the new incoming samples have a sample size $m = n_{j+1} - n_j$ that converges to infinity, but instead, we assume that K converges to infinity, while $m = n_j - n_{j-1}$ is constant. This represents a realistic situation in which finite chunks of data regularly come in resulting in eventual large sample sizes n_j for j large. Asymptotics for an estimator can be characterized by $K \to \infty$, which is equivalent with $n = n_K = Km$ converging to infinity.

In the context of the Big Data era n might be of the order 10^9 making it computationally intractable to recompute a fixed estimator based on each updated sample with sample size n_j as j increases from 1 to $K = 10^9/m$, even when utilizing super-computers. Instead, optimally speaking from a computational point of view, we want an estimator in which one updates a current estimator based on new computations that are bounded by (e.g.) $O(m^p)$-calculations, for some fixed p. As a consequence, such a procedure for obtaining a sequence of K estimators only involves $KO(m^p)$ number of calculations and such a procedure will be able to update the estimators as fast as data comes in so that the user will always have available the most up to date estimator. On the other hand, a procedure that recalculates the estimator for each updated sample will take roughly $KO(n^p)$ number of calculations, making it completely intractable for such large sample sizes. As we will see, bounding the new computations by $O(m^p)$ for some fixed p is possible but might come at a cost of how data adaptive the estimator can be and might thus hurt fundamental statistical properties such as consistency and asymptotic normality. Therefore, we will provide a flexible enough framework so that these choices can be made based on a user supplied trade-off, but either way, the resulting online estimators will be computationally superior relative to a non online estimator.

Let $\Psi : \mathcal{M} \rightarrow \mathbb{R}^d$ be a Euclidean target parameter mapping of interest so that $\Psi(P_0)$ denotes the desired estimand we want to learn from the data. Suppose that $\Psi(P) = \Psi_1(Q(P))$ for some parameter mapping Ψ_1 and parameter $P \rightarrow Q(P)$ on $\mathcal{M}$, so that $\Psi(P)$ only depends on P through a smaller part $Q(P)$ of the data generating distribution P, which we often refer to as the relevant part of the data generating distribution. For notational convenience, recognizing the abuse of notation, we will denote Ψ_1 with Ψ as well. Let $(Q, O) \rightarrow L(Q)(O)$ be a loss-function for Q_0 so that $Q_0 = \arg\min_{Q \in Q\mathcal{M}} P_0 L(Q)$, where we use the notation $Pf \equiv \int f(o)dP(o)$, and $Q(\mathcal{M}) \equiv \{Q(P) : P \in \mathcal{M}\}$ for the parameter space of Q.

Assume that Ψ is pathwise differentiable at P for each $P \in \mathcal{M}$ and let $O \rightarrow D^*(P)(O)$ be the efficient influence curve of $\Psi : \mathcal{M} \rightarrow \mathbb{R}^d$ at $P \in \mathcal{M}$, which is defined as the canonical gradient of the pathwise derivative along parametric paths through P: for any one dimensional path $\{P(\epsilon) : \epsilon\} \subset \mathcal{M}$ through P with score $S = \frac{d}{d\epsilon} \log P(\epsilon)\big|_{\epsilon=0}$ at $\epsilon = 0$ we have

$$\frac{d}{d\epsilon}\Psi(P(\epsilon))\bigg|_{\epsilon=0} = PD^*(P)S.$$

This canonical gradient is uniquely defined as the only gradient $D(P)$ (i.e, each component is an element of $L_0^2(P)$ and $PD^*(P)S = PD(P)S$ for all scores S) whose components are also an element of the so called tangent space $T(P)$ defined as the closure of the linear span of all the scores generated by the class of parametric paths.

Suppose that the canonical gradient $D^*(P)$ only depends on P through $Q(P)$ and a nuisance parameter $G(P)$ defined as functions of P on the model $\mathcal{M}$: to emphasize this we will use the notation $D^*(P) = D^*(Q(P), G(P))$ for some $(Q, G) \rightarrow D^*(Q, G)$. We remind the reader that the efficient influence curve is a crucial mathematical element one calculates from the model $\mathcal{M}$ and the definition of target parameter mapping $\Psi : \mathcal{M} \rightarrow \mathbb{R}^d$ that defines efficiency of an estimator of ψ_0: an estimator ψ_n is an asymptotically efficient estimator of ψ_0 if and only if

$$\psi_n - \psi_0 = (P_n - P_0)D^*(P_0) + o_P(1/\sqrt{n}) = \frac{1}{n}\sum_{i=1}^{n} D^*(P_0)(O_i) + o_P(1/\sqrt{n}).$$

In words, this states that an estimator is efficient at $P_0 \in \mathcal{M}$ if and only if the estimator is asymptotically linear at P_0 with influence curve equal to the efficient influence curve $D^*(P_0)$. By the convolution theorem, an efficient estimator is the asymptotically best estimator among the class of all regular estimators (and is itself a regular estimator).

Let $R(P, P_0)$ be defined by

$$P_0 D^*(P) = \Psi(P_0) - \Psi(P) + R(P, P_0),$$

where, by the fact that $D^*(P)$ is the canonical gradient of the pathwise derivative so that $(P_0 - P)D^*(P)$ can be interpreted as a first order expansion of $\Psi(P_0) - \Psi(P)$, $R(P, P_0)$ is a second order remainder that can be explicitly determined given Ψ and D^*. Equivalently, in terms of $D^*(P) = D^*(Q, G)$ and $\Psi(P) = \Psi(Q)$, we have

$$P_0 D^*(Q, G) = \Psi(Q_0) - \Psi(Q) + R(Q, G, Q_0, G_0) \tag{2}$$

for a specified second order term $R()$.

Let $\mathbf{O}_k = (O_{n_{k-1}+1}, \ldots, O_{n_k})$ represent the $m = n_k - n_{k-1}$ observations making up batch k, $k = 1, 2, \ldots, K$, where $n_0 = 0$. For notational convenience, we define

$$D_k^*(P)(\mathbf{O}_k) \equiv \frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k} D^*(P)(O_i).$$

Before we proceed with presenting our proposed online estimators of ψ_0 in the next sections, let's first formally define what we mean with an online estimator.

Definition 1 *An online estimator of a parameter $\psi_0 = \Psi(P_0)$ based on a sequence of batches $\mathbf{O}_1, \mathbf{O}_2, \ldots$ is a sequence of estimators $(\psi_k : k = 1, \ldots)$ with ψ_k being an estimator based on $\mathbf{O}_1, \ldots, \mathbf{O}_k$ satisfying the following property: there exist certain functions f_1 and f_2, and a sequence of estimators $(\eta_k : k = 1, \ldots)$ with $\eta_k = f_2(\mathbf{O}_k, \eta_{k-1})$, so that*

$$\psi_k = f_1(\mathbf{O}_k, \eta_{k-1}),\ k = 1, \ldots.$$

This definition can be applied to our target parameter ψ_0, but also to define an online estimator of (Q_0, G_0). Different online estimators can differ drastically in their required memory storage and computational speed by using an online estimator η_k that grows in dimension with k versus an η_k that has a fixed dimension in k.

Let $((Q_k, G_k) : k = 1, \ldots,)$ be an online estimator of (Q_0, G_0). For example, this might be estimators using a stochastic gradient descent algorithm based on a very high dimensional parametric model, but more flexible online estimators are presented in Section 6. In the next sections we will propose a variety of online estimators of ψ_0 that map this online estimator $((Q_k, G_k) : k = 1, \ldots)$ into an online estimator ψ_k of ψ_0, so that ψ_k is only a function of $(Q_{k-1}, G_{k-1}, \psi_{k-1})$, and possibly a few more online low-dimensional statistics, and the new batch $\mathbf{O}_k$.

A crucial ingredient in the analysis of our proposed online estimators is the following identity that is an immediate consequence of (9):

$$P_{0,k} D_k^*(Q_{k-1}, G_{k-1}) = \Psi(Q_0) - \Psi(Q_{k-1}) + R(Q_{k-1}, G_{k-1}, Q_0, G_0), \ k = 1, \ldots, K, \tag{3}$$

where we used the notation $P_{0,k} f(\mathbf{O}_k) = \int f(\mathbf{O}_k) dP_{0,k}(\mathbf{O}_k)$ and $dP_{0,k}(\mathbf{O}_k) = \prod_{i=n_{k-1}+1}^{n_k} dP_0(O_i)$ is the probability distribution of $\mathbf{O}_k$ implied by the common probability distribution P_0 and the fact that all O_i are independent. Note that we also have $P_{0,k} D_k^*(Q_{k-1}, G_{k-1}) = E_0(D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) \mid \mathcal{F}_{k-1})$ is the conditional expectation of the random variable $D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k)$ (a function of $\mathbf{O}_1, \ldots, \mathbf{O}_k$), given $\mathcal{F}(k-1) \equiv (\mathbf{O}_1, \ldots, \mathbf{O}_{k-1})$.

We will assume that an initial estimator $Q_{k=0}, G_{k=0}$ is given, so that the online procedure can be initiated with this choice. In practice this might be an estimator based on an initial batch that is further ignored in our definition on the online estimator. However, one could also simply define $(Q_{k=0}, G_{k=0}) = (Q_1, G_1)$, i.e., as the online estimator based on the first batch, since this choice does not affect the asymptotics (i.e., it only affects the impact of the first $n_1 = m$ observations in the online estimator which is asymptotically negligible as $K \to \infty$).

3 Online one-step estimator

Define the following online one-step estimator of ψ_0:

$$\psi_K = \frac{1}{K} \sum_{k=1}^{K} \left\{ \Psi(Q_{k-1}) + D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) \right\}. \tag{4}$$

Equivalently, this can be calculated in an online manner as follows:

$$\psi_K = \frac{K-1}{K}\psi_{K-1} + \frac{1}{K}\{\Psi(Q_{K-1}) + D_K^*(Q_{K-1}, G_{K-1})(\mathbf{O}_K)\}.$$

We have the following theorem.

Theorem 1
Definitions: *Let* $\bar{M}(K) = \sum_{k=1}^K M_k$*, where* $M_k = D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - P_{0,k}D_k^*(Q_{k-1}, G_{k-1})$*. We have that* $(\bar{M}(k) : k = 1, \ldots)$ *is a discrete martingale w.r.t.* $\mathcal{F}_k = (\mathbf{O}_1, \ldots, \mathbf{O}_k)$*: that is,* $E_0(\bar{M}(K) \mid \mathcal{F}(k)) = \bar{M}(k)$ *for* $k \leq K$*. Let*

$$\Sigma_k^2 \equiv E_0 M_k^2 \equiv E_0 M_k M_k^\top,$$

and

$$\Sigma^2(K) \equiv \frac{1}{K}\sum_{k=1}^K \Sigma_k^2.$$

Define also

$$W^2(K) = \frac{1}{K}\sum_{k=1}^K E_0(M_k^2 \mid \mathcal{F}_{k-1}) = \frac{1}{K}\sum_{k=1}^K P_{0,k}M_k^2.$$

Define

$$\bar{R}(K) \equiv \frac{1}{K}\sum_{k=1}^K R_0(Q_{k-1}, G_{k-1}, Q_0, G_0).$$

We have the following expansion for the online one-step estimator:

$$\psi_K - \psi_0 = \frac{\bar{M}(K)}{K} + \bar{R}(K).$$

Assumptions: *We make the following assumptions*

- *For some* $M < \infty$ $\max_k \mid D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) \mid < M < \infty$ *with probability 1;*
- $\bar{R}(K) = o_P(1/\sqrt{K})$*;*
- $\liminf_{K\to\infty} \lambda\Sigma^2(K)\lambda > 0$ *for all* λ*, or that*

  $\Sigma^2 = \lim_{k\to\infty} \Sigma(k)^2$ *exists and is a positive definite covariance matrix;*

- $W^2(K) - \Sigma^2(K) \rightarrow_{K\rightarrow\infty} 0$ *in probability, or, equivalently*

$$\frac{1}{K}\sum_{k=1}^{K} P_{0,k}D_k^*(Q_{k-1},G_{k-1}) - E_0\frac{1}{K}\sum_{k=1}^{K} P_{0,k}D_k^*(Q_{k-1},G_{k-1}) \rightarrow 0 \qquad (5)$$

in probability as $K \rightarrow \infty$.

Conclusion: *Then,*

$$\Sigma(K)^{-1}\frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0,I),\ \textit{as } K\rightarrow\infty,$$

and, if $\Sigma^2(K) \rightarrow \Sigma^2$, *as* $K \rightarrow \infty$, *for some positive definite covariance matrix* Σ^2, *then*

$$\frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0,\Sigma^2),\ \textit{as } K\rightarrow\infty.$$

This implies:

$$\sqrt{Km}(\psi_K - \psi_0) \Rightarrow_D N(0,\Sigma^2/m)\ \textit{as } K\rightarrow\infty.$$

We have that $\Sigma^2/m = P_0D^*(Q_0,G_0)^2$ *is the efficiency bound, which proves that* ψ_K *is an asymptotically efficient estimator of* ψ_0.

Consistent estimation of asymptotic variance: *Finally, consider the following estimator of* $\Sigma^2(K)$*:*

$$\hat{\Sigma}^2(K) = \frac{1}{K}\sum_{k=1}^{K}\{D_k^*(Q_{k-1},G_{k-1}(\mathbf{O}_k) - \bar{D}_K\}^2,$$

where $\bar{D}_K = \frac{1}{K}\sum_{k=1}^{K} D_k^*(Q_{k-1},G_{k-1})(\mathbf{O}_k)$. *We have* $\hat{\Sigma}^2(K) - \Sigma^2(K) \rightarrow 0$ *in probability as* $K \rightarrow \infty$, *and if* Σ^2 *exists, then we also have* $\hat{\Sigma}^2(K) \rightarrow \Sigma^2$ *in probability as* $K \rightarrow \infty$.

We note that this theorem proves that the online one-step estimator is asymptotically efficient without any restriction on how data adaptive Q_k, G_k can be. We only need that these estimators are consistent in a weak sense (for the purpose of the martingale weak convergence), and they need to converge at a good enough rate so that $\bar{R}(K) = o_P(1/\sqrt{K})$. That is, contrary to the analysis of a regular one-step estimator or TMLE that rely on $D_k^*(Q_k,G_k)$ to fall in a Donsker class with probability tending to one, this theorem does not require any Donsker class conditions. This is due to the fact that $\bar{M}(K)$ is a martingale process, which relies on the fact that we estimate the unknown (Q_0,G_0) in $D_k^*(Q_0,G_0)(\mathbf{O}_k)$ with estimators that are based on the past $\mathbf{O}_1,\ldots,\mathbf{O}_{k-1}$. In that sense, this online TMLE achieves the same as the cross-validated TMLE that uses internal sample splitting to remove the Donsker class condition.

3.1 Central limit theorem for discrete martingales

Our theorem relies on establishing weak convergence of the process $(\bar{M}(K)/\sqrt{K} : K)$ as $K \to \infty$. For that purpose we apply a central limit theorem for discrete martingales. An example of such a theorem is given in Sen and Singer (1993), resulting in Theorem 17 in van der Laan (2008). In our context this Theorem 17 translates into the following one.

Theorem 2 *Let* $\bar{M}(K) = \sum_{k=1}^{K} M_k$, $M_k = (M_{k1}, \ldots, M_{kd})$, $E_0(M_k \mid \mathcal{F}_{k-1}) = 0$, *where* $\mathcal{F}_k = (\mathbf{O}_1, \ldots, \mathbf{O}_k)$. *In our case,* $M_k = D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - P_{0,k} D_k^*(Q_{k-1}, G_{k-1})$.
Definitions: *Let*

$$\Sigma_k^2 \equiv E_0 M_k^2 \equiv E_0 M_k M_k^\top,$$

and

$$V_k^2 \equiv E_0 \left(M_k^2 \mid \mathcal{F}_{k-1} \right) = P_{0,k} M_k^2.$$

Let

$$\Sigma^2(K) \equiv \frac{1}{K} \sum_{k=1}^{K} \Sigma_k^2 = E_0 \frac{1}{K} \sum_{k=1}^{K} P_{0,k} M_k^2$$

and

$$W^2(K) \equiv \frac{1}{K} \sum_{k=1}^{K} V_k^2 = \frac{1}{K} \sum_{k=1}^{K} P_{0,k} M_k^2.$$

Assumptions: *Assume that for some* $M < \infty$ $\max_k \mid D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) \mid M < \infty$ *with probability 1;* $\liminf \lambda \Sigma(k)^2 \lambda > 0$ *for all* λ *(or that* $\Sigma^2 = \lim_{k \to \infty} \Sigma(k)^2$ *exists and is a positive definite covariance matrix); and that component wise*

$$\frac{1}{K} \sum_{k=1}^{K} P_{0,k} D_k^*(Q_{k-1}, G_{k-1}) - E_0 \frac{1}{K} \sum_{k=1}^{K} P_{0,k} D_k^*(Q_{k-1}, G_{k-1}) \to 0 \tag{6}$$

in probability as $K \to \infty$.
Conclusion: *Then,*

$$\Sigma(K)^{-1} \frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0, I), \text{ as } K \to \infty,$$

and, if $\Sigma^2(K) \to \Sigma^2$, *as* $K \to \infty$, *for some positive definite covariance matrix* Σ^2, *then*

$$\frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0, \Sigma^2), \text{ as } K \to \infty.$$

3.2 Proof of Theorem 1:

We have

$$\begin{aligned}\psi_K &= \frac{1}{K}\sum_{k=1}^{K}\{\Psi(Q_{k-1}) + D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - P_{0,k}D_k^*(Q_{k-1}, G_{k-1}))\} \\ &\quad + \frac{1}{K}\sum_{k=1}^{K} P_{0,k}D_k^*(Q_{k-1}, G_{k-1}).\end{aligned}$$

By identity (10), we have

$$\frac{1}{K}\sum_{k=1}^{K} P_{0,k}D_k^*(Q_{k-1}, G_{k-1}) = \psi_0 - \frac{1}{K}\sum_{k=1}^{K}\Psi(Q_{k-1}) + \frac{1}{K}\sum_{k=1}^{K} R_0(Q_{k-1}, G_{k-1}, Q_0, G_0).$$

Substitution of this in the last expression yields now

$$\psi_K - \psi_0 = \frac{\bar{M}(K)}{K} + \bar{R}(K),$$

where

$$\begin{aligned}\bar{M}(K) &= \sum_{k=1}^{K}\{D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - P_{0,k}D_k^*(Q_{k-1}, G_{k-1}))\} \\ \bar{R}(K) &= \frac{1}{K}\sum_{k=1}^{K} R_0(Q_{k-1}, G_{k-1}, Q_0, G_0).\end{aligned}$$

We assumed that $\bar{R}(K) = o_P(1/\sqrt{K})$ (or equivalently, $\bar{R}(K) = o_P(1/\sqrt{n})$. We now note that $\bar{M}(K) = \sum_{k=1}^{K} M_k$, where $E_0(M_k \mid \mathbf{O}_1, \ldots, \mathbf{O}_{k-1}) = 0$. Thus, $E_0(\bar{M}(K) \mid \mathbf{O}_1, \ldots, \mathbf{O}_k) = \bar{M}(k)$, which proves that $(\bar{M}(k) : k)$ is a discrete martingale process. Application of Theorem 2 to $\bar{M}(K)$ establishes the conclusions of Theorem 1, and, in particular, $\frac{\bar{M}(K)}{\sqrt{K}}$ converges to $N(0, \Sigma^2)$. Finally, the fact that $\Sigma^2/m = P_0 D^*(Q_0, G_0)^2$ is easily verified. The consistency of the estimator of $\Sigma^2(K)$ is a consequence of Sen and Singer (1993), formally presented by Theorem 3 below. This completes the proof. □

3.3 Estimation of limit covariance matrix of multivariate martingale sum.

Note that the natural estimator of $\Sigma^2(K) = E_0\frac{1}{K}\sum_{k=1}^{K} P_{0,k}\{D_k^*(Q_0, G_0) - P_{0,k}D_k^*(Q_0, G_0)\}^2$ is given by

$$\hat{\Sigma}^2(K) = \frac{1}{K}\sum_{k=1}^{K}\{D_k^*(Q_{k-1}, G_{k-1}(\mathbf{O}_k) - \bar{D}_K\}^2,$$

where $\bar{D}_K = \frac{1}{K}\sum_{k=1}^{K} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k)$. The following results proves that this estimator of the covariance matrix of the multivariate margingale $1/\sqrt{K}\sum_{k=1}^{K} M_k$ is indeed asymptotically consistent, under the same conditions we needed for establishing its weak convergence. The following theorem is taken from Sen and Singer (1993) and corresponds with Theorem 18 in van der Laan (2008).

Theorem 3 *Under the conditions stated in Theorem 2, we have that*

$$\hat{\Sigma}^2(K) - \Sigma(K)^2 \to 0 \textit{ in probability, as } K \to \infty,$$

and, if $\Sigma^2(K) \to \Sigma^2$, as $K \to \infty$, for a positive definite matrix Σ^2, then this also implies $\hat{\Sigma}^2(K) \to \Sigma$ in probability, as $K \to \infty$.

4 Online targeted one-step estimator

Consider a least favorable submodel $\{Q(\epsilon \mid G) : \epsilon\}$ of $\{Q(P) : P \in \mathcal{M}\}$ so that the linear span of $\frac{d}{d\epsilon} L(Q(\epsilon \mid G))(O)\big|_{\epsilon=0}$ contains the linear span of the components of $D^*(Q, G)(O)$. Suppose now that, given the initial online estimator $(Q_k, G_k : k = 1, \ldots)$, we construct an online estimator $(\epsilon_k : k)$ so that

$$\frac{1}{K}\sum_{k=1}^{K} D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k) = o_P(1/\sqrt{K}), \tag{7}$$

where

$$Q_k^* = Q_k(\epsilon_k \mid G_k).$$

We refer to (7) as the online-efficient influence curve estimation equation, in order to contrast it with the efficient influence curve estimating equation $1/K\sum_{k=1}^{K} D_k^*(Q_K^*, G_K)(\mathbf{O}_k) = 0$ solved by the regular TMLE.

To be general, let $((Q_k^*, G_k) : k)$ be any online estimator of (Q_0, G_0) satisfying (7). In this case, our one-step estimator (4) reduces to

$$\Psi_k = \frac{1}{K}\sum_{k=1}^{K} \Psi(Q_{k-1}^*) + \frac{1}{K}\sum_{k=1}^{K} D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k).$$

Since the last term is now $o_P(1/\sqrt{K})$, this estimator is asymptotically equivalent with

$$\Psi_k^* \equiv \frac{1}{K}\sum_{k=1}^{K} \Psi(Q_{k-1}^*).$$

The latter type estimators we will refer to as an online TMLE. Such an estimator is attractive by being an average of substitution estimators, which naturally respect the global constraints of the model $\mathcal{M}$ and target parameter. The same theorem 1 applies to such an estimator.

4.1 A particular online TMLE if batch size converges to infinity at appropriate rate

The challenge is now how to construct such an online targeted estimator $((Q_k^*, G_k) : k)$ that solves the martingale efficient estimating equation (7) up till an $o_P(1/\sqrt{K})$-term. One possible approach is to define, for each k, Q_k^* as the TMLE based on initial estimator (Q_k, G_k), least favorable submodel $\{Q_k(\epsilon \mid G_k) : \epsilon\}$ only using the k-th batch $\mathbf{O}_k = (O_i : i = (k-1)m+1, \ldots, km)$. That is, we define

$$\epsilon_k = \arg\min_{\epsilon} P_{k,m} L(Q_k(\epsilon \mid G_k)) = \arg\min_{\epsilon} \frac{1}{m} \sum_{i=n_{k-1}+1}^{n_k} L(Q_k(\epsilon \mid G_k))(O_i),$$

and define the update $Q_k^1 = Q_k(\epsilon_k \mid G_k)$, where $P_{k,m}$ is the empirical distribution of the k-th batch $\mathbf{O}_k$. This updating process is iterated till convergence and the final update is denoted with Q_k^*. Such an online TMLE solves $D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k) = 0$, for all k, so that, certainly, (7) is solved exactly (i.e. $o_P(1/\sqrt{K})$ replaced by 0). Unfortunately, we now run into the issue that this ϵ_k will have bias $O(1/m)$ (like any parametric MLE), and that will not converge to zero as $k \to \infty$, so that Q_K^* will not be a consistent estimator of Q_0, even when Q_K is consistent. This online TMLE can be shown to be asymptotically efficient if one is willing to let $m = m(n)$ to converge to infinity at a not too slow rate, and it appears to be a good and practically interesting estimator in finite samples for relatively small batch sizes m, based on our simulation studies.

In the next subsections we present an online ϵ_k^0, analogue to a second order stochastic gradient descent algorithm, that could be used to target the online one step estimator (i.e., set $\epsilon_k = \epsilon_k^0$ above). It remains to be seen if this online estimator does the job (7), which we plan to address in future research. Instead of diving into the analysis of this targeted one-step estimator, we will propose an alternative online TMLE in the next section which will be formally analyzed.

4.2 Stochastic gradient descent (SGD) algorithm for a parametric model

Below we present an algorithm that defines an initial estimator ϵ_k^0 of ϵ_{0k} which happens to be an analogue of the stochastic gradient descent algorithm for approximating the MLE for a parametric model. In this subsection we review this quickly, before presenting this initial online estimator ϵ_k^0.

Let $\theta_0 = \arg\max_\theta P_0 \log p_\theta$ for a parametric model $\{p_\theta : \theta\}$ of densities. Let $S(\theta)(O) = \frac{d}{d\theta}\log p_\theta(O)$ and let $S^*(\theta)(O) = -\{P_0 \frac{d}{d\theta} S(\theta)\}^{-1} S(\theta)(O)$ be the efficient score. Under weak regularity conditions, we have that $P_0 S(\theta_0) = P_0 S^*(\theta_0) = 0$. Finally, let $S_k^*(\theta)(\mathbf{O}_k) = \frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k} S^*(\theta)(O_i)$. We now define the following analogue of our online one-step estimator with the twist that the initial external online estimator is replaced by θ_k itself:

$$\theta_K = \frac{1}{K}\sum_{k=1}^{K}\theta_{k-1} + \frac{1}{K}\sum_{k=1}^{K} S_k^*(\theta_{k-1})(\mathbf{O}_k).$$

This iterative algorithm for evaluating θ_K needs to be initiated with an initial $\theta_{k=0}$.

By using the above expression for θ_K and θ_{K-1} and taking the difference, it follows that this algorithm can actually be written as:

$$\theta_K = \theta_{K-1} + \frac{1}{K} S_K^*(\theta_{K-1})(\mathbf{O}_K).$$

This is exactly the second order stochastic gradient descent algorithm, standardizing the gradient by minus the inverse of its derivative, as in the Newton-Raphson algorithm, which is a particular class of stochastic gradient descent algorithms in the literature. As discussed in Section 1, results in the literature show that this algorithm has good convergence properties (although straightforward modifications might be employed to guarantee convergence), and that it approximates the MLE $\theta_{mK} = \arg\max_\theta P_{n_K} \log p_\theta$ up till an asymptotically negligible $o_P(1/\sqrt{n_K})$.

4.3 Using second order SGD to construct initial online estimator of ϵ_{0k}.

We consider the following algorithm for constructing an initial online estimator ϵ_K^0, which is the recursive analogue of our online one-step estimator in which the initial online estimator is replaced by its previous realizations, just as we did in the previous subsection. Start with an $\epsilon^0_{k=0}$ such as one that minimizes $\epsilon \rightarrow D^*_1(Q_{k=0}(\epsilon \mid G_{k=0}), G_{k=0})(\mathbf{O}_1)$, and iteratively compute $\epsilon^0_1, \ldots$, as follows:

$$\epsilon^0_K = \frac{1}{K}\sum_{k=1}^{K} \epsilon^0_{k-1} + \frac{1}{K}\sum_{k=1}^{K} c^{-1}_{k-1} D^*_k(Q_{k-1}(\epsilon^0_{k-1} \mid G_{k-1}), G_{k-1})(\mathbf{O}_k),$$

where $c_{k-1} \equiv \frac{d}{d\epsilon^0_{k-1}} \Psi(Q_{k-1}(\epsilon^0_{k-1} \mid G_{k-1}))$.

Subtracting the right-hand sides for these two equations for ϵ^0_K and ϵ^0_{K-1} shows that this algorithm can also be formulated as:

$$\epsilon^0_K = \epsilon^0_{K-1} + \frac{1}{K} c^{-1}_{K-1} D^*_K(Q_{K-1}(\epsilon^0_{K-1} \mid G_{K-1}), G_{K-1})(\mathbf{O}_K).$$

It remains to be formally established if this online estimator ϵ^0_k will indeed satisfy (7).

5 Online TMLE

In this section we will define an online estimator ϵ^*_k, which takes as input the online estimators Q_k, G_k and an initial ϵ^0_k, which yields an asymptotically efficient substitution estimator $\Psi(Q_K(\epsilon^*_K \mid G_K))$ of ψ_0, which we call an online TMLE, where, as with the regular TMLE, $\{Q_k(\epsilon \mid G_k) : \epsilon\}$ is the least favorable submodel through Q_k at $\epsilon = 0$ so that the linear span of $\frac{d}{d\epsilon} L(Q_k(\epsilon \mid G_k))(O)\Big|_{\epsilon=0}$ contains $D^*(Q_k, G_k)(O)$.

Let ϵ^0_k be an initial estimator of ϵ_{0k}, where ϵ_{0k} is defined as the solution of

$$f_k(\epsilon) \equiv \Psi(Q_k(\epsilon \mid G_k)) = \psi_0.$$

If Q_k is a consistent estimator of Q_0, a requirement for our efficiency theorem above for the one-step estimator, and below for the online TMLE, then we could set $\epsilon^0_k = 0$.

Even though our theorems present conditions for asymptotic efficiency that rely on consistency of Q_k, in the case that $R_0(Q, G, Q_0, G_0)$ satisfies a so called double robustness structure (i.e. $R_0(Q, G, Q_0, G_0) = 0$ if either $Q = Q_0$ or $G = G_0$), the online one-step estimator can remain asymptotically linear if either Q_k or G_k is consistent, but not necessarily both, and in such a case ϵ_{0k} will generally not converge to zero. Such a more general theorem for the one-step estimator would be a completely analogue to these types of theorems presented for the regular TMLE, and will not be repeated here.

Therefore, we also want to discuss initial estimators ϵ^0_k that are not relying on consistency of Q_k. For example, if ψ^1_k is our online-one step estimator, then we could set ϵ_k^0 equal to the solution of $f_k(\epsilon) = \psi_k^1$, but this would rely on ψ_k^1 to be an element of the parameter space of ψ_0. If ψ_k^1 is not an element of the parameter space one could replace it by its projection $\tilde{\psi}_k^1$ on the parameter space and select ϵ_k^0 as an approximate solution of $f_k(\epsilon) = \tilde{\psi}_k^1 + o_P(1/\sqrt{k})$ (no need to exactly solve it). Most importantly, in the last subsection of the previous section we provide a stochastic gradient descent algorithm generating a sequence of ϵ_k^0 to target the online one-step estimator and one could select this one as the initial estimator of ϵ_{0k}.

We use this initial estimator ϵ_k^0 of ϵ_{0k} to construct a linear approximation of $f_k(\epsilon)$ in a neighborhood of ϵ_{0k}:

$$\tilde{f}_k(\epsilon) = f_k(\epsilon_k^0) + \frac{d}{d\epsilon_k^0} f_k(\epsilon_k^0)(\epsilon - \epsilon_k^0).$$

We will use the short-hand notation $c_k \equiv \frac{d}{d\epsilon_k^0} f_k(\epsilon_k^0)$. Under weak regularity conditions, we have

$$f_k(\epsilon) = \tilde{f}_k(\epsilon) + o_P(\parallel \epsilon - \epsilon_k^0 \parallel),$$

where the remainder is a second order term. We approximate the inverse of the non-linear function f_k on a neighborhood of ϵ_{0k} with the inverse of the linear approximation $\tilde{f}_k$ of f_k, and the latter inverse is given by:

$$\tilde{f}_k^{-1}(\psi) = \epsilon_k^0 + c_k^{-1}(\psi - f_k(\epsilon_k^0)).$$

Under weak regularity conditions, we will also have that

$$R_0(\epsilon_k^0, \epsilon_{0k}) \equiv \tilde{f}_k^{-1}(\psi_0) - f_k^{-1}(\psi_0) = o(\parallel \epsilon_k^0 - \epsilon_{0k} \parallel), \tag{8}$$

which is the crucial result our proposed algorithm for ϵ_K^* below relies upon. In other words, the inverse of our linear approximation approximates the inverse of f_k at ψ_0 up till a second order term.

We now define the following online estimator ϵ_K^* of ϵ_{0K}:

$$\epsilon_K^* = \frac{1}{K}\sum_{k=1}^{K} c_K^{-1}\Psi(Q_{k-1}) + \frac{1}{K}\sum_{k=1}^{K} \tilde{f}_K^{-1} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k).$$

This defines our online TMLE $\Psi(Q_K(\epsilon_K^* \mid G_K))$ of ψ_0. We have the following theorem for this online estimator ϵ_K^*:

Theorem 4 *Consider the definitions $\bar{M}(K)$ and $\bar{R}(K)$ in Theorem 1. Recall the above definitions of f_k, $\tilde{f}_k$, c_K, $R_0(\epsilon_k^0, \epsilon_{0k}) \equiv \tilde{f}_k^{-1}(\psi_0) - f_k^{-1}(\psi_0)$ and ϵ_{0k} defined by $\Psi(Q_k(\epsilon \mid G_k)) = \psi_0$. Let $c_{0K} \equiv \frac{d}{d\epsilon_{0K}}\Psi(Q_K(\epsilon_{0K} \mid G_K))$.*

We have the following expansion:

$$\epsilon_K^* - \epsilon_{0K} = c_K^{-1} \frac{\bar{M}(K)}{K} + c_K^{-1} \bar{R}(K) + R_0(\epsilon_K^0, \epsilon_{0K}).$$

Assumptions: *Assume the same conditions as in Theorem 1: i.e.,* $\bar{R}(K) = o_P(1/\sqrt{K})$*, and the martingale consistency conditions on* (Q_k, G_k) *so that* $\bar{M}(K)/\sqrt{K}$ *converges to optimal multivariate normal mean zero distributed* $Z \sim N(0, \Sigma^2)$ *as specified in Theorem 1. In addition, assume* $R_0(\epsilon_K^0, \epsilon_{0K}) = o_P(1/\sqrt{K})$*,* $\frac{d}{d\epsilon}\Psi(Q_K(\epsilon))$ *is continuous at* $\epsilon = \epsilon_{0K}$*, the inverse* c_{0K}^{-1} *of this derivative at* ϵ_{0K} *has a bounded norm uniformly in* K*, so that, in particular,*

$$c_{0K} c_K^{-1} \rightarrow 1 \text{ in probabilty as } K \rightarrow \infty.$$

Conclusion: *Then,*

$$\epsilon_K^* - \epsilon_{0K} = c_{0K}^{-1} \frac{\bar{M}(K)}{K} + o_P(1/\sqrt{K}),$$

and $\sqrt{K}(\epsilon_K^* - \epsilon_{0K}) - c_{0K}^{-1} Z \rightarrow 0$ *in probability.*

An immediate corollary of Theorem 4 is the following Theorem providing asymptotic efficiency of the online TMLE of ψ_0 (a consequence of the delta-method).

Theorem 5 *Under the same conditions as in Theorem 4, we have*

$$\Psi(Q_K(\epsilon_K^* \mid G_K)) - \Psi(Q_0) = \frac{\bar{M}(K)}{K} + o_P(1/\sqrt{K}),$$

and thereby $\sqrt{K}(\Psi(Q_K(\epsilon_K^* \mid G_K)) - \Psi(Q_0)) \Rightarrow_d Z$*, where* $Z \sim N(0, \Sigma^2)$ *is the optimal normal mean zero distribution defined in Theorem 1. Thus* $\Psi(Q_K(\epsilon_K^* \mid G_K))$ *is an asymptotically efficient estimator of* ψ_0*.*

The proofs of the two theorems above are presented in the Appendix.

6 Online super learning.

Our efficient online estimators of the target parameter ψ_0 rely on initial online estimators of nuisance parameters (Q_0, G_0). We could use the online SGD algorithms for parametric models, where one could decide to use a very high dimensional parametric model in order to make it flexible enough to provide a good approximation of the true (Q_0, G_0). The choice of parametric model will still be a delicate issue and might have a significant effect on the final estimator. In particular, by selecting the dimension of the parametric model a priori, one might be forced to select the dimension much too large for the first part of the data, and eventually the model might simply not be adaptive enough. Therefore, it would be helpful to allow for a more data adaptive approach for online estimation. For that purpose we propose the following general template for construction of an online data adaptive estimator of (Q_0, G_0). For the sake of explanation, we can focus on Q_0, since Q_0 and G_0 are estimated separately. Recall that $(O, Q) \rightarrow L(Q)(O)$ is a loss function so that $Q_0 = \arg\min_Q P_0 L(Q)$, where we minimize over the parameter space $Q(\mathcal{M}) \equiv \{Q(P) : P \in \mathcal{M}\}$ for Q_0.

For a given $Q^0 \in Q(\mathcal{M})$ and a measure of its precision n^0, let $\hat{Q}_j(\cdot \mid Q^0, n^0) : \mathcal{M}_{NP} \rightarrow Q(\mathcal{M})$, $j = 1, \ldots, J$, be candidate estimators of Q_0 that can be applied to any empirical distribution (i.e., member of the nonparametric model $\mathcal{M}_{NP}$) of a given data set, such as the empirical distribution $P_{k,m}$ of the m observations in $\mathbf{O}_k$ at the k-th stage. It is assumed that these candidate estimators take as input, beyond the next batch of data, an off-set Q^0 and a measure of its precision (i.e., the sample size it was based upon), which will play the role of the current value Q_{k-1} of the estimator right before the k-th batch of data. For example,

$$\hat{Q}_j(P_m \mid Q^0, n^0) = \frac{n^0 Q^0 + m\hat{Q}_{j1}(P_m \mid Q^0)}{n^0 + m},$$

or

$$\hat{Q}_j(P_m \mid Q^0, n^0) = \frac{1}{1 + \exp\left(-\{n^0 \mathrm{logit} Q^0 + m \mathrm{logit} \hat{Q}_{j1}(P_m \mid Q^0)\}/(n^0 + m)\right)}.$$

Here $\hat{Q}_j(P_m \mid Q^0, n^0)$ is some function of Q^0, n^0 and an estimator $\hat{Q}_{j1}(P_m \mid Q^0)$ that aims to learn a deviation from the offset Q^0 based on P_m. For example, $\hat{Q}_{j1}(P_m \mid Q^0)$ might be a fit $Q_{\beta_m}(Q^0)$ of a parametric model $\{Q_\beta(Q^0) : \beta\}$ through Q^0 at $\beta = 0$, where $\beta_m = \arg\min_\beta P_m L(Q_\beta(Q^0))$.

Given a parametric function f_α indexed by a finite dimensional parameter α, let $\hat{Q}_\alpha(P_m \mid Q^0, n^0) = f_\alpha((\hat{Q}_j(P_m \mid Q^0, n^0) : j = 1, \ldots, J))$ be a new estimator that combines the candidate estimators $(\hat{Q}_j(P_m \mid Q^0, n^0) : j = 1, \ldots, J)$ into a new estimator $\hat{Q}_\alpha(P_m \mid Q^0, n^0)$. For example, we might have that $\hat{Q}_\alpha(P_m \mid Q^0, n^0) = \sum_{j=1}^J \alpha_j \hat{Q}_j(P_m \mid Q^0, n^0)$ is a weighted average of the candidate estimators. For a given cross-validation scheme $B_m \in \{0, 1\}^m$ and empirical distribution P_m of m observations, let P^0_{m,B_m}, P^1_{m,B_m} be the empirical distributions of the training sample $\{O_i : B_m(i) = 0\}$ and validation sample $\{O_i : B_m(i) = 1\}$, respectively, which partition the sample of m observations. We select α with the cross-validation selector:

$$\alpha(P_m \mid Q^0, n^0) \equiv \arg\min_{\alpha} E_{B_m} P^1_{m,B_m} L(\hat{Q}_\alpha(P^0_{m,B_m} \mid Q^0, n^0)).$$

We can now define our offset-specific super-learner estimator based on P_m:

$$\hat{Q}(P_m \mid Q^0, n^0) = \hat{Q}_{\alpha(P_m|Q^0,n^0)}(P_m \mid Q^0, n^0).$$

This off-set specific super-learner defines now the following online super-learner of Q_0. Let $P_{k,m}$ be the empirical distribution of the k-th batch $\mathbf{O}_k = \{O_i : i = n_{k-1}+1, \ldots, n_k\}$, where $n_{k+1} - n_k = m$. Define $Q_{k=0} = 0$. Let $Q_1 = \hat{Q}(P_{1,m} \mid Q_{k=0}, 0)$, and, iterate this sequential updating process with $Q_{k+1} = \hat{Q}(P_{k+1,m} \mid Q_k, k)$, $k = 2, \ldots, K-1$.

Note that $(Q_k : k = 1, 2, \ldots)$ is indeed an online estimator since at stage k, the computation of Q_k is only based on the new batch of data $\mathbf{O}_k$, and the current value Q_{k-1}. The heuristic of the online estimator is as follows. We use a super-learner supported by the theoretical properties of cross-validation for each batch $\mathbf{O}_k$ (van der Laan and Dudoit, 2003; van der Vaart et al., 2006; van der Laan et al., 2006, 2007; Polley et al., 2012), but we only use this new batch of data to learn the update relative to the current estimate Q_{k-1}. If m is reasonably large, then this allows one to use a flexible approach to learn this update and the cross-validation controls the overfitting on the k-th batch. The theoretical properties of the cross-validation selector at each stage follows from the application of the oracle inequality for the cross-validation selector, applied conditional on the independent past data $(\mathbf{O}_1, \ldots, \mathbf{O}_{k-1})$ (or more succinctly, one only conditions on Q_{k-1}). However, the online estimator will still respect the fact that Q_{k-1} was based on $(k-1)m$ observations while the new batch only consists of m observations: each of the candidate estimators in the super-learner weights the offset Q_{k-1} with $k-1$ and the update based on $\mathbf{O}_k$ with 1. Clearly, the variance of Q_K corresponds with the variance of K independent estimators and will thus behave well. The main concern should therefore be the bias of the online estimator. However, by only using the new m observations to fit a residual of the type $Q_0 - Q_{k-1}$, conditional on Q_{k-1}, one would expect that there will be bias reduction at each step. For example, if $Q_0 = E_0(Y \mid X)$, then this is comparable with using a super-learner to fit the regression of $Y(Q_{k-1}) = Y - Q_{k-1}$ on X: so one uses the new batch $\mathbf{O}_k$ to fit

$$E_0(Y - Q_{k-1}(X) \mid X) = E_0(Y \mid X) - Q_{k-1}(X) = Q_0(X) - Q_{k-1}(X).$$

Such type of updating algorithms have been proven to be successful in bias reduction as well: e.g. SGD for fitting a parametric model has been shown to have bias $O(1/n)$, just like the MLE, and the iterative TMLE-algorithm iteratively removes bias by iteratively using maximum likelihood estimation along a parametric submodel through the current estimator, and has been shown to result in semi parametric efficient estimators.

We can apply the same online super-learner approach to obtain an online estimator $(G_k : k = 1, \ldots)$ for G_0.

7 Example: Online efficient one-step estimator for a parametric model.

Let $\{p_\theta : \theta\}$ be a parametric working model of densities indexed by finite dimensional parameter θ. The target parameter mapping is $\Psi : \mathcal{M} \to \mathbb{R}^d$ defined by $\Psi(P) = \arg\max_\theta P \log p_\theta$ and the model $\mathcal{M}$ is the nonparametric model. For this target parameter and model, the efficient influence curve at P is given by $D^*(P)(O) = -c_\psi(P)^{-1}S(\psi)(O)$, where $S(\psi) = \frac{d}{d\psi} \log p_\psi$ is the score at ψ, and $c_\psi(P) = \frac{d}{d\psi} PS(\psi)$, where $\psi = \Psi(P)$. For convenience, we use the notation $c_0 = c_{\psi_0}(P_0)$ for the true normalizing matrix. We note that $D^*(P)$ only depends on P through $\Psi(P)$ and $c_{\Psi(P)}(P)$, so that we can also represent $D^*(P)$ as $D^*(\Psi, c_\Psi)$, and $D^*(P_0) = D^*(\psi_0, c_0)$.

Let $D^*_k(\psi, c) = \frac{1}{m} \sum_{i=n_{k-1}+1}^{n_k} D^*(\psi, c)(O_i)$ be efficient score for the k-th batch $\mathbf{O}_k$. Consider the following online estimator ψ^*_k for estimation of $\psi_0 = \arg\max_\theta P_0 \log p_\theta$, or equivalently under weak regularity conditions, the solution of $P_0 S(\theta) = P_0 D^*(\theta, c_\theta) = 0$.

- Let ψ^0_k be an initial online estimator that converges to ψ_0 at a rate faster than $n^{-1/4}$ (e.g., the standard stochastic gradient descent algorithm converges at rate $1/\sqrt{n}$).
- Let $c^0_k = c_{\psi^0_k, k}$ be an online estimator of $c_{\psi^0_k}(P_0)$.
- Define
$$\psi^*_K = \frac{1}{K} \sum_{k=1}^{K} \psi^0_{k-1} + \frac{1}{K} \sum_{k=1}^{K} D^*_k(\psi^0_{k-1}, c^0_{k-1})(\mathbf{O}_k).$$

In the next theorem we use the notation $P_0 D^2 = P_0 D D^\top$ for a vector value function $O \to D(O)$.

Theorem 6
Definitions: *Let* $R_0(\psi, c, \psi_0, c_0)$ *be defined by* $P_0 D^*(\psi, c_\psi) = \psi_0 - \psi + R_0(\psi, c_\psi, \psi_0, c_0)$ *(note* $R_0(\psi, c_\psi, \psi_0, c_0)$ *is a second order term). Let* ψ_k^0, c_k^0 *be online estimators of* ψ_0, c_0, *respectively, and define*

$$\psi_K^* = \frac{1}{K}\sum_{k=1}^{K}\psi_{k-1}^0 + \frac{1}{K}\sum_{k=1}^{K} D_k^*(\psi_{k-1}^0, c_{k-1}^0)(\mathbf{O}_k).$$

Assumptions: *Assume that*

$$\bar{R}(K) \equiv \frac{1}{K}\sum_{k=1}^{K} R_0(\psi_k^0, c_k^0, \psi_0, c_0) = o_P(1/\sqrt{K}),$$

and assume that ψ_K^0 *is consistent for* ψ_0 *so that the martingale*

$$\bar{M}(K)/\sqrt{K} \equiv \frac{1}{\sqrt{K}}\sum_{k=1}^{K}\{D_k^*(\psi_{k-1}^0, c_{k-1}^0)(\mathbf{O}_k) - P_{0,k}D_k^*(\psi_{k-1}^0, c_{k-1}^0)\}$$

converges weakly to a random variable Z *with normal distribution, as* $K \to \infty$, *equal to the normal limit distribution of*

$$\frac{1}{\sqrt{K}}\sum_{k=1}^{K} D_k^*(\psi_0, c_0)(\mathbf{O}_k).$$

Conclusion: *We have* $(\psi_K^* - \psi_0) = \bar{M}(K)/K + \bar{R}(K)$, *and thus* $\sqrt{K}(\psi_K^* - \psi_0) \Rightarrow_d Z$ *as well, as* $K \to \infty$, *or equivalently,* $\sqrt{n}(\psi_{K(n)}^* - \psi_0) \Rightarrow_d N(0, \Sigma_0 = P_0 D^*(\psi_0, c_0)^2)$. *In particular,* $\Psi_{K(n)}^*$ *is an asymptotically efficient estimator of* ψ_0.

This is an immediate corollary of Theorem 1.

8 Example: Online one-step efficient estimation of the counterfactual mean.

Let $O = (W, A, Y) \sim P_0$, and $\mathcal{M}$ is the nonparametric model. The target parameter $\Psi : \mathcal{M} \to \mathbb{R}$ is defined as $\Psi(P) = E_P E_P(Y \mid A = 1, W)$. Under a causal model (Neyman, 1990; Rubin, 1974, 2006; Holland, 1986; Robins, 1987a,b; Pearl, 2009)), this parameter equals the counterfactual mean under treatment. Note that $\Psi(P)$ only depends on P through $Q(P) = (Q_W(P), \bar{Q}(P))$, where $Q_W(w) = P(W \leq w)$, $\bar{Q}(A, W) = E_P(Y \mid A, W)$. Therefore, we will also denote the target parameter with $\Psi(Q)$. An online estimator of $Q_{W,0}$ is easy to construct by just augmenting the current sample with the new batch and mapping the current empirical distribution into the empirical distribution of the complete sample. However, such an online estimator requires a memory that grows linearly with sample size n_K. Therefore, we first recognize that $\Psi(Q)$ only depends on Q_W through $Q_W\bar{Q} = \int \bar{Q}(1, W)dQ_W(W)$: i..e., we only need to know the mean of $\bar{Q}$. We will denote this parameter with $\bar{Q}_W = Q_W\bar{Q}$, so that we will redefine $Q = (\bar{Q}, \bar{Q}_W)$, and $\Psi(Q) = \Psi(\bar{Q}, \bar{Q}_W) = \bar{Q}_W$.

The efficient influence curve is given by (see, e.g., van der Laan and Robins (2003); Rose and van der Laan (2011))

$$D^*(Q, G)(O) = \frac{A}{G(1 \mid W)}(Y - \bar{Q}(1, W)) + \bar{Q}(1, W) - \Psi(Q),$$

where $G(1 \mid W) = P(A = 1 \mid W)$. Note that $D^*(Q, G) = D^*(\bar{Q}, \bar{Q}_W, G)$.

We have

$$P_0 D^*(Q, G) = \Psi(Q_0) - \Psi(Q) + R_0(\bar{Q}, G, \bar{Q}_0, G_0),$$

where

$$R_0(\bar{Q}, G, \bar{Q}_0, G_0) = E_{P_0}(\bar{Q} - \bar{Q}_0)(1, W)\frac{G - G_0}{G}(1 \mid W).$$

This also defines the efficient influence curve $D_k^*(Q, G)(\mathbf{O}_k)$ for the k-th batch $\mathbf{O}_k$, and the corresponding identity

$$P_{0,k} D_k^*(Q, G) = \Psi(Q_0) - \Psi(Q) + R_0(\bar{Q}, G, \bar{Q}_0, G_0).$$

8.1 Online one-step estimator

Online estimator of $Q_0 = (\bar{Q}_0, \bar{Q}_{W0})$ and G_0: We assume that the online estimator $(\bar{Q}_k : k = 0, \ldots, K)$ and $(G_k : k = 0, \ldots, K)$ is given, being one of the methods we discussed.

We now present a corresponding online estimator $(\bar{Q}_{W,k} : k = 0, \ldots, K)$. Given an online estimator $(\bar{Q}_k : k = 0, \ldots, K)$, $\bar{Q}_{W,k=0} = 0$, we define for $k = 1, \ldots, K$

$$\bar{Q}_{W,k} = \frac{k-1}{k}\bar{Q}_{W,k-1} + \frac{1}{k}\frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k} \bar{Q}_{k-1}(1, W_i),$$

which defines the online estimator $(\bar{Q}_{W,k} : k = 0, \ldots, K)$. Note that

$$\bar{Q}_{W,K} = \frac{1}{K} \sum_{k=1}^{K} \frac{1}{m} \sum_{i=n_{k-1}+1}^{n_k} \bar{Q}_{k-1}(1, W_i).$$

This defines now an online estimator $(Q_k = (\bar{Q}_k, \bar{Q}_{W,k}) : k = 0, \ldots, K)$ of $Q_0 = (\bar{Q}_0, \bar{Q}_{W0})$.
Online one-step estimator of ψ_0: Note that $\Psi(Q_{k-1}) = \bar{Q}_{W,k-1}$. The online one-step estimator of ψ_0 is defined as:

$$\psi_K = \frac{1}{K} \sum_{k=1}^{K} \Psi(Q_{k-1}) + \frac{1}{K} \sum_{k=1}^{K} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k).$$

Of course, this can be calculated in an online manner, since

$$\psi_K = \frac{K-1}{K} \psi_{K-1} + \frac{1}{K} \{\Psi(Q_{K-1}) + D_K^*(Q_{K-1}, G_{K-1})(\mathbf{O}_K)\}.$$

Asymptotics: By Theorem 1 we have that, if $\bar{R}(K) = \frac{1}{K} \sum_{k=1}^{K} R_0(Q_{k-1}, G_{k-1}, Q_0, G_0) = o_P(1/\sqrt{K})$, and Q_K, G_K are consistent estimators of Q_0, G_0 in the variance norm sense as defined by the martingale conditions in Theorem 1, then $\sqrt{Km}(\psi_K - \psi_0) \rightarrow_d N(0, P_0 D^*(Q_0, G_0)^2)$ as $K \rightarrow \infty$, and thus that ψ_K is asymptotically efficient.

Due to the double robustness structure of $R_0()$, if Q_K converges to a possibly misspecified Q, but G_K is consistent, under additional conditions, by approximating $\bar{R}_K$ in first order with a martingale process $1/K \sum_{k=1}^{K} D_{1k}(Q_{k-1}, G_{k-1})(\mathbf{O}_k)$ plus a $o_P(1/\sqrt{K})$-term, ψ_K remains asymptotically normal with asymptotic variance $P_0\{(D_k^* + D_{1k})(Q, G_0)\}^2$.

9 Example: Online TMLE of counterfactual mean

Recall our general definition of the online TMLE, since we will follow here the same template for defining this online TMLE and, in particular, ϵ_k^*.

The online TMLE is defined in terms of a least favorable model used to update an initial online estimator. Given the initial online estimator $(G_k, \bar{Q}_k : k)$, we select the following least favorable submodel through the initial online estimator $\bar{Q}_k$

$$\text{Logit}\bar{Q}_k(\epsilon \mid G_k) = \text{Logit}\bar{Q}_k + \epsilon H(G_k),$$

where $H(G_k)(A,W) = \frac{A}{G_k(A|W)}$. In this section the online estimator of $\bar{Q}_{W,0} = Q_{W,0}\bar{Q}_0$ is given by $\bar{Q}_{W,k} = \frac{1}{mk}\sum_{i=1}^{km}\bar{Q}_k(1,W_i)$, i.e., it is simply the empirical mean of all elements $(\bar{Q}_k(1,W_i) : i = 1,\ldots,mk)$. Equivalently, one could define $Q_k = (Q_{W,k},\bar{Q}_k)$, where $Q_{W,k}$ is the empirical distribution of $(W_i : i = 1,\ldots,mk)$.

Below, we will construct an online estimator $(\epsilon_k^* : k = 0,\ldots,K)$. This now defines the online estimators $(\bar{Q}_k^* = \bar{Q}_k(\epsilon_k^* \mid G_k) : k = 0,\ldots,K)$ and corresponding

$$\Psi(Q_k^*) = \Psi(Q_{W,k},\bar{Q}_k(\epsilon_k^*)) = \frac{1}{mk}\sum_{i=1}^{mk}\bar{Q}_k(\epsilon_k^*)(1,W_i).$$

In the next subsection we define the online estimator ϵ_k^*.

9.1 Online estimator ϵ_k^*.

For now, let an initial online estimator $(\epsilon_k^0 : k = 0,\ldots,K)$ be given (e.g., if Q_k is consistent, then we can select $\epsilon_k^0 = 0$), but below we will show how to obtain this with the second order SGD algorithm. As above, this defines an online estimator $Q_k^0 = (\bar{Q}_k(\epsilon_k^0),\bar{Q}_{W,k}^0)$, where

$$\bar{Q}_{W,k}^0 = \frac{1}{mk}\sum_{i=1}^{mk}\bar{Q}_k(\epsilon_k^0)(1,W_i).$$

Let ϵ_0 represent the limit of ϵ_k^0 as $k \to \infty$, defined by $\Psi(Q(\epsilon \mid G)) = \psi_0$ where (Q,G) represent the limit of (Q_k,G_k). Note that $\epsilon_0 = 0$ if Q_k is consistent for Q_0.

Online estimator of standardizing constant: We also need an online estimator of $c_0 \equiv \frac{d}{d\epsilon_0}\Psi(Q_{W,0},\bar{Q}(\epsilon_0 \mid G))$. We have $\Psi(Q_{W,0},\bar{Q}(\epsilon_0 \mid G)) = E_{0,W}\bar{Q}(\epsilon_0)(1,W)$. Thus, $\frac{d}{d\epsilon_0}E_{0,W}\bar{Q}(\epsilon_0)(1,W) = E_{0,W}\frac{d}{d\epsilon_0}\bar{Q}(\epsilon_0)(1,W)$. We have

$$\frac{d}{d\epsilon_0}\bar{Q}(\epsilon_0)(1,W) = -\frac{\bar{Q}(\epsilon_0)(1-\bar{Q}(\epsilon_0))(1,W)}{G(1 \mid W)}.$$

Thus, we have

$$c_0 = -E_{0,W}\frac{\bar{Q}(\epsilon_0)(1-\bar{Q}(\epsilon_0))(1,W)}{G(1 \mid W)}.$$

Given the online estimators Q_k, G_k, ϵ_0^k, the online estimator of c_0 is naturally defined as follows:

$$c_K \equiv -\frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{i=n_{k-1}}^{n_k}\frac{\bar{Q}_{k-1}(\epsilon_{k-1}^0)(1-\bar{Q}_{k-1}(\epsilon_{k-1}^0))}{G_{k-1}}(1,W_i).$$

As shown above, this can be iteratively computed showing it is indeed an online estimator:

$$c_k = \frac{k-1}{k}c_{k-1} + \frac{1}{k}\frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k}\frac{\bar{Q}_{k-1}(\epsilon_{k-1}^0)(1-\bar{Q}_{k-1}(\epsilon_{k-1}^0))}{G_{k-1}}(1,W_i).$$

Combining with online estimator ϵ_k^0: We now show that we can also use the SGD to obtain the online estimator ϵ_k^0. Let $\bar{Q}_{k-1}, \bar{Q}_{W,k-1}, G_{k-1}, \epsilon_{k-1}^0, c_{k-1}$ be given. We can now immediately compute c_k as a function of $(c_{k-1}, \bar{Q}_{k-1}(\epsilon_{k-1}^0), G_{k-1})$ and $\mathbf{O}_k$. We can also immediately compute $(\bar{Q}_k, G_k)$ and thereby the empirical average $\bar{Q}_{W,k}$ of $\bar{Q}_k$. The next ϵ_k^0 is defined by the SGD:

$$\epsilon_k^0 = \epsilon_{k-1}^0 + \frac{1}{k}c_{k-1}^{-1}D_k^*(Q_{k-1}(\epsilon_{k-1}^0 \mid G_{k-1}), G_{k-1})(\mathbf{O}_k).$$

This shows that we have mapped $(\bar{Q}_{k-1}, \bar{Q}_{W,k-1}, G_{k-1}, \epsilon_{k-1}^0, c_{k-1})$ and $\mathbf{O}_k$ into $(\bar{Q}_k, \bar{Q}_{W,k}, G_k, \epsilon_k^0, c_k)$. This shows that we have defined an online estimator $(\bar{Q}_k, \bar{Q}_{W,k}, G_k, \epsilon_k^0, c_k : k = 0, \ldots, K)$.

The inverse of $\tilde{f}_k$ is given by:

$$\tilde{f}_k^{-1}(\psi) = \epsilon_k^0 + c_k^{-1}(\psi - \Psi(Q_k^0)).$$

This demonstrates that we have an online estimator $(\bar{Q}_k, \bar{Q}_{W,k}, G_k, \epsilon_k^0, c_k, \tilde{f}_k^{-1} : k = 0, \ldots, K)$.

We now define the following online estimator ϵ_K^* of ϵ_{K0}:

$$\epsilon_K^* = \frac{1}{K}\sum_{k=1}^{K}c_K^{-1}\Psi(Q_{k-1}) + \frac{1}{K}\sum_{k=1}^{K}\tilde{f}_K^{-1}D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k).$$

We note that

$$\begin{aligned}\epsilon_K^* - \epsilon_{K-1}^* &= \frac{(K-1)c_K^{-1} - Kc_{K-1}^{-1}}{K(K-1)}\sum_{k=1}^{K-1}\Psi(Q_{k-1}) + \frac{c_K^{-1}}{K}\Psi(Q_{K-1}) \\ &+\frac{(K-1)\tilde{f}_K^{-1} - K\tilde{f}_{K-1}^{-1}}{K(K-1)}\sum_{k=1}^{K-1}D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) + \frac{\tilde{f}_K^{-1}}{K}D_K^*(Q_{K-1}, G_{K-1})(\mathbf{O}_K).\end{aligned}$$

This shows that for an explicitly defined function f, we have

$$\epsilon_K^* = f(K, \epsilon_{K-1}^*, c_K, c_{K-1}, \bar{\Psi}(K-1), \Psi(Q_{K-1}), \tilde{f}_K, \tilde{f}_{K-1}, \bar{D}(K-1), Q_{K-1}, G_{K-1}, \mathbf{O}_K),$$

where $\bar{\Psi}(K-1) = \sum_{k=1}^{K-1} \Psi(Q_{k-1})$ and $\bar{D}(K-1) = \sum_{k=1}^{K-1} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k)$. This proves that ϵ_K^* is indeed also an online estimator.

9.2 Asymptotics of online TMLE

We can apply Theorem 5 to establish asymptotic efficiency of this online TMLE $\psi_K^* = \Psi(Q_K(\epsilon_K^*))$ of ψ_0 under appropriate conditions. Specifically, if $\bar{R}(K) = \frac{1}{K}\sum_{k=1}^K R_0(Q_{k-1}, G_{k-1}, Q_0, G_0) = o_P(1/\sqrt{K})$, and Q_K, G_K are consistent estimators of Q_0, G_0 in the variance norm sense as defined by the martingale conditions in Theorem 1, then $\sqrt{Km}(\psi_K^* - \psi_0) \rightarrow_d N(0, P_0 D^*(Q_0, G_0)^2)$ as $K \rightarrow \infty$, and thus that ψ_K^* is asymptotically efficient.

Due to the double robustness structure of $R_0()$, if Q_K^* converges to a possibly misspecified Q, but G_K is consistent, under additional conditions, by approximating $\bar{R}_K$ in first order with a martingale process $1/K \sum_{k=1}^K D_{1k}(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k)$, ψ_K^* remains asymptotically normal with asymptotic variance $P_0\{(D_k^* + D_{1k})(Q, G_0)\}^2$.

Remark Suppose that $\bar{Q}_k$ is defined by a finite dimensional vector of parameters. In that case, the computational burden of computing the update $\bar{Q}_k(\epsilon_k^*)$ is controlled by the fixed dimension. However, computing $\bar{Q}_k(\epsilon_k^*)$ at each of the mk observations will require mk evaluations which thus increases linearly in k. However, one might only carry out the evaluation of the empirical mean of $\bar{Q}_k(\epsilon_k^*)$ at special occasions, while also running the online-one step estimator of ψ_0. In that case, the computational burden of the online TMLE is still controlled. On the other hand, in order to be nonparametrically consistent we should let the dimension of $\bar{Q}_k$ grow with sample size. There are probably sensible strategies to make the dimension of $\bar{Q}_k$ grow with sample size at a much slower rate than mk so that the computational burden is reasonably controlled. Either way, this online TMLE ($\psi_k^* : k = 1, \ldots, K$) is still much faster than recomputing once a TMLE at the last stage K, and that does not even take into account that one might really need to have an estimator at each stage $k = 1, \ldots, K$.

9.3 Targeted online one-step estimator of counterfactual mean

We now construct a targeted version of the online one-step estimator that might be asymptotically equivalent with $\frac{1}{K}\sum_{k=1}^K \Psi(Q_{k-1}(\epsilon_k^0))$, although that would rely on (7) to hold, which has not been formally established. This targeted one-step estimator is as fast to compute as the online one-step estimator, and has therefore computational advantages relative to the online TMLE above that is slowed down by the evaluation of $\Psi()$ at an Q_k^*, while it might still be a substitution estimator.

Online estimator of $Q_0 = (\bar{Q}_0, \bar{Q}_{W0})$ and G_0: We assume that the online estimator $(\bar{Q}_k : k = 0, \ldots, K)$ and $(G_k : k = 0, \ldots, K)$ is given, being one of the methods we discussed.

Online estimator Q_k^*, G_k: Let $\bar{Q}_{k-1}, \bar{Q}_{W,k-1}, G_{k-1}, \epsilon_{k-1}^0, c_{k-1}$ be given. This defines $\bar{Q}_{k-1}^* = \bar{Q}_{k-1}(\epsilon_{k-1}^0)$ and the corresponding

$$\bar{Q}_{W,k-1}^* = \frac{k-1}{k}\bar{Q}_{W,k-2}^* + \frac{1}{k}\frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k}\bar{Q}_{k-1}^*(1, W_i),$$

which defines $Q_{k-1}^* = (\bar{Q}_{k-1}^*, \bar{Q}_{W,k-1}^*)$. Note that for any K:

$$\bar{Q}_{W,K}^* = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k}\bar{Q}_{k-1}^*(1, W_i).$$

Given the online estimator c_k above, it follows that we can now immediately computer c_k as a function of $(c_{k-1}, \bar{Q}_{k-1}(\epsilon_{k-1}^0), G_{k-1})$ and $\mathbf{O}_k$:

$$c_k = \frac{k-1}{k}c_{k-1} + \frac{1}{k}\frac{1}{m}\sum_{i=n_{k-1}+1}^{n_k}\frac{\bar{Q}_{k-1}(\epsilon_{k-1}^0)(1-\bar{Q}_{k-1}(\epsilon_{k-1}^0))}{G_{k-1}}(1, W_i).$$

The next ϵ_k^0 is now defined by the SGD:

$$\epsilon_k^0 = \epsilon_{k-1}^0 + \frac{1}{k}c_{k-1}^{-1}D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k).$$

We can also compute the updates $(\bar{Q}_k, G_k)$ in an online manner. This shows that we have mapped $(\bar{Q}_{k-1}, \bar{Q}_{W,k-1}, G_{k-1}, \epsilon_{k-1}^0, c_{k-1}, \bar{Q}_{k-1}^*, \bar{Q}_{W,k-1}^*)$ and $\mathbf{O}_k$ into $(\bar{Q}_k, \bar{Q}_{W,k}, G_k, \epsilon_k^0, c_k, \bar{Q}_k^*, \bar{Q}_{W,k}^*)$. This shows that we have defined an online estimator $(\bar{Q}_k, \bar{Q}_{W,k}, G_k, \epsilon_k^0, c_k, \bar{Q}_k^*, \bar{Q}_{W,k}^* : k = 0, \ldots, K)$, and thus in particular an online estimator (Q_k^*, G_k).

Targeted online one-step estimator of ψ_0: Note that $\Psi(Q_{k-1}^*) = \bar{Q}_{W,k-1}^*$. The online one-step estimator of ψ_0 is defined as:

$$\psi_K^* = \frac{1}{K}\sum_{k=1}^{K}\Psi(Q_{k-1}^*) + \frac{1}{K}\sum_{k=1}^{K}D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k).$$

Of course, this can be calculated in an online manner, since

$$\psi_K^* = \frac{K-1}{K}\psi_{K-1}^* + \frac{1}{K}\left\{\Psi(Q_{K-1}^*) + D_K^*(Q_{K-1}^*, G_{K-1})(\mathbf{O}_K)\right\}.$$

Due to using a targeted estimator Q_k^*, we wonder if $\frac{1}{K}\sum_{k=1}^K D_k^*(Q_{k-1}^*, G_{k-1})(\mathbf{O}_k) = o_P(1/\sqrt{K})$, in which case this targeted online one-step estimator is asymptotically equivalent with the substitution estimator $\frac{1}{K}\sum_{k=1}^K \Psi(Q_{k-1}^*)$. Either way, we expect that the empirical average correction will be relatively small so that the targeting will still have robustified the online one step estimator.

The asymptotic efficiency of this online one step estimator is established by Theorem 1 again.

10 Generalization to online learning based on an adaptive group sequential design

In this section we generalize our formulation and online estimators to group sequential adaptive designs, thereby allowing that the data generating distribution changes over time and is possibly data adaptively adjusted in response to the observed past (van der Laan, 2008; Chambaz and van der Laan, 2011a,b; Chambaz et al., 2014).

10.1 Formulation of the online estimation problem for a group sequential adaptive design

Let $O_1, \ldots, O_n$ be such that O_i, given $O_1, \ldots, O_{i-1}$, has probability distribution $P_{\theta_0, g_{0,i}}$, $i = 1, \ldots, n$, where $g_{0,i}$ is a known function of $\bar{O}(i-1)$, and $\theta_0 \in \Theta$ for some parameter space Θ. Let $\mathcal{M}_i = \{P_{\theta, g_{0,i}} : \theta \in \Theta\}$ be the model for $P_{\theta_0, g_{0,i}}$, $i = 1, \ldots, n$. Suppose that our target quantity of interest is $\Psi^F(\theta_0)$ for some $\Psi^F : \Theta \to \mathbb{R}$, and that it is identifiable through a pathwise differentiable parameter $\Psi_i : \mathcal{M}_i \to \mathbb{R}$ in the sense that $\Psi_i(P_{\theta, g_{0,i}}) = \Psi^F(\theta)$ for each $\theta \in \Theta$. Let $O_i \to D^*(\theta, g_{0,i})(O_i)$ be the efficient influence curve of $\Psi_i : \mathcal{M}_i \to \mathbb{R}$ at $P_{\theta, g_{0,i}}$.

Suppose that $\Psi_i(P_{\theta, g_{0,i}})$ only depends on θ through a parameter $Q(\theta)$ of θ, and assume that $D^*(\theta, g_{0,i})(O_i)$ only depends on θ through $(Q(\theta), G(\theta))$ for some nuisance parameter $G(\theta)$ of θ, so that we can also denote the target parameter and efficient influence curve with $\Psi_i(Q)$ and $D^*(Q, G, g_{0,i})$, respectively. Let $(Q, O) \to L(Q, g_{0,i})(O)$ be a loss-function for $Q_0 = Q(\theta_0)$ based on sampling from $P_{\theta_0,g_{0,i}}$ in the sense that

$$Q_0 = \arg\min_Q P_{\theta_0,g_{0,i}} L(Q, g_{0,i}) = \arg\min_Q \int L(Q, g_{0,i})(o) dP_{\theta_0,g_{0,i}}(o),$$

where we minimize over the parameter space $\{Q(\theta) : \theta \in \Theta\}$ of Q_0.

Our goal is to construct online estimator ψ_0^F so that

$$\psi_n - \psi_0^F = \frac{1}{n} \sum_{i=1}^{n} D^*(Q_0, G_0, g_{0,i})(O_i) + o_P(1/\sqrt{n}),$$

so that the martingale CLT establishes weak convergence of $\sqrt{n}(\psi_n - \psi_0)$ to $N(0, \sigma^2)$, under appropriate conditions. In particular, if $g_{0,i}$ converges to a fixed parameter g_0 as $i \to \infty$, then $\sigma^2 = P_{\theta_0,g_0} D^*(Q_0, G_0, g_0)^2$. That is, the estimator ψ_n achieves the same asymptotic performance as an efficient estimator based on i.i.d. sampling from P_{θ_0,g_0}. In our work referenced above on targeted adaptive designs, we have referred to $g_{0,i}$ as the targeted adaptive design targeting the (possibly) optimal g_0.

Let $R_0(\theta, \theta_0, g_{0,i})$ be defined by

$$P_{\theta_0,g_{0,i}} D^*(\theta, g_{0,i}) = \Psi^F(\theta_0) - \Psi^F(\theta) + R_0(\theta, \theta_0, g_{0,i}),$$

where $R_0(\theta, \theta_0, g_{0,i})$ is a second order remainder that can be explicitly determined given Ψ_i and $D^*(\theta, g_{0,i})$. Equivalently, in terms of $D^*(Q, G, g_{0,i})$ and $\Psi_i(Q)$, we have

$$P_{\theta_0,g_{0,i}} D^*(Q, G, g_{0,i}) = \Psi^F(Q_0) - \Psi^F(Q) + R_0(Q, G, Q_0, G_0, g_{0,i}). \qquad (9)$$

Let $\mathbf{O}_k = (O_{n_{k-1}+1}, \ldots, O_{n_k})$ represent the $m = n_k - n_{k-1}$ observations making up batch k, $k = 1, 2, \ldots, K$, where $n_0 = 0$, and assume $g_{0,i} = \mathbf{g}_{0,k}$ is constant across i in the k-th batch. For notational convenience, we define

$$D_k^*(Q, G, \mathbf{g}_{0,k})(\mathbf{O}_k) \equiv \frac{1}{m} \sum_{i=n_{k-1}+1}^{n_k} D^*(Q, G, g_{0,i})(O_i),$$

where we could also replace $g_{0,i}$ by $\mathbf{g}_{0,k}$ in the last expression.

Let $((Q_k, G_k) : k = 1, \ldots,)$ be an online estimator of (Q_0, G_0). A crucial ingredient in the analysis of our proposed online estimators is the following identity that is an immediate consequence of (9):

$$P_{0,k} D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k}) = \Psi^F(Q_0) - \Psi^F(Q_{k-1}) + R_0(Q_{k-1}, G_{k-1}, Q_0, G_0, \mathbf{g}_{0,k}), \qquad (10)$$

for $k = 1, \ldots, K$. Here we used the notation $P_{0,k} f(\mathbf{O}_k) = \int f(\mathbf{O}_k) dP_{0,k}(\mathbf{O}_k)$ and $dP_{0,k}(\mathbf{O}_k) = \prod_{i=n_{k-1}+1}^{n_k} dP_{\theta_0, g_{0,i}} O_i)$ is the probability density of $\mathbf{O}_k$, conditional on $\bar{O}(n_{k-1})$. Note that we also have

$$P_{0,k} D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k}) = E_0(D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})(\mathbf{O}_k) \mid \mathcal{F}_{k-1})$$

is the conditional expectation of the random variable $D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})(\mathbf{O}_k)$ (a function of $\mathbf{O}_1, \ldots, \mathbf{O}_k$), given $\mathcal{F}(k-1) \equiv (\mathbf{O}_1, \ldots, \mathbf{O}_{k-1})$.

We will assume that an initial estimator $(Q_{k=0}, G_{k=0})$ is given, so that the online procedure can be initiated with this choice. In practice this might be an estimator based on an initial batch that is further ignored in our definition on the online estimator. However, one could also simply define $(Q_{k=0}, G_{k=0}) = (Q_1, G_1)$, since this choice does not affect the asymptotics.

10.2 A few examples of data adaptive designs.

Let's consider two concrete examples that demonstrate this type of setting. Firstly, consider the case that $X_i = (W_i, Y_{1,i}, Y_{0,i}) \sim_{iid} P_{X,0}$ consists of baseline covariates W_i, and two treatment specific counterfactual outcomes $Y_{1,i}, Y_{0,i}$ and let $O_i = (W_i, A_i, Y_i \equiv Y_{A_i,i})$ be the missing data structure with censoring indicator equal to the binary treatment A_i. One could imagine a group sequential randomized trial in which we keep recruiting subjects over time by sampling them from a particular fixed population, but that the randomization probabilities for A_i change over time in response to past data $\bar{O}(i-1)$: given $\bar{O}(i-1)$, $P(A_i = 1 \mid W_i) = \Pi_{i,0}(1 \mid W_i)$ is known (Bai et al. (2002); Andersen et al. (1994); Flournoy and Rosenberger (1995); Hu and Rosenberger (2000); Rosenberger (1996); Rosenberger et al. (1997); Rosenberger and Grill (1997); Rosenberger and Shiram (1997); Tamura et al. (1994); Wei (1979); Wei and Durham (1978); Wei et al. (1990); Zelen (1969); Cheng and Shen (2005)). . In this case, the conditional probability distribution of O_i, given $\bar{O}(i-1)$, $P_{P_{X,0}, \Pi_{i,0}}$ is indexed by the unknown common distribution $P_{X,0}$ and the known treatment assignment mechanism $\Pi_{0,i}$. The full-data target parameter might be defined as $\Psi^F : \mathcal{M}^F \rightarrow \mathbb{R}$ with $\Psi^F(P_{X,0}) = E_{P_{X,0}}(Y_1 - Y_0)$, where $\mathcal{M}^F$ is the nonparametric model for $P_{X,0}$. This full-data target parameter

$$\psi_0^F = \Psi^F(P_{X,0}) = \Psi(P_{0,i}) \equiv E_{P_{0,i}}\{E_{P_{0,i}}(Y_i \mid A_i = 1, W_i) - E_{P_{0,i}}(Y_i \mid A_i = 0, W_i)\}$$

is identifiable from $P_{0,i} = P_{P_{X,0}, \Pi_i}$ for each i if $0 < \Pi_{i,0}(1 \mid W_i) < 1$.

As a second example, suppose now that $X_i = (W_i, A_i, Y_i)$ corresponds with an observation in an observational study so that we cannot view the conditional distribution of A_i, given $(W_i, Y_{1,i}, Y_{0,i})$, as known, but we still assume that X_i

are independent and follow a common distribution $P_{X,0}$. Assume that we observe on the i-th unit $O_i = (W_i, A_i, \Delta_i, \Delta_i Y_i)$, a missing data structure on X_i with missing indicator Δ_i indicating if the outcome is observed. Assume that, conditional on $\bar{O}(i-1)$, the missingness mechanism $\Pi_{i,0}(\Delta_i \mid X_i)$ is known. For example, given $\bar{O}(i-1)$, one uses case-control sampling so that $\Pi_{i,0}(\Delta_i \mid X_i)$ only depends on X_i through Y_i, W_i according to a known probability distribution, learned from past data. As a consequence, the conditional distribution of O_i, given $\bar{O}(i-1)$, $P_{P_{X,0},\Pi_{i,0}}$ is now again indexed by the common distribution $P_{X,0}$ of X_i and this known missingness mechanism $\Pi_{i,0}$. The target quantity is defined as $\Psi^F(P_{X,0}) = E_{P_{X,0}}\{E_{P_{X,0}}(Y \mid A=1, W) - E_{P_{X,0}}(Y \mid A=0, W)\}$ and we would pose a certain model $\mathcal{M}^F$ for the probability distribution of (W, A, Y) such as the nonparametric model if nothing is known about the conditional distribution of A, given W. So in this example we have that X_i does include censoring/treatment variables that are not within the control of the experimenter.

10.3 Online one-step estimator

Define the following one-step estimator of ψ_0^F:

$$\psi_K^F = \frac{1}{K}\sum_{k=1}^{K}\left\{\Psi^F(Q_{k-1}) + D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})(\mathbf{O}_k)\right\}. \tag{11}$$

We have the following theorem.

Theorem 7
Definitions: *Let* $\bar{M}(K) = \sum_{k=1}^{K} M_k$*, where* $M_k = D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})(\mathbf{O}_k) - P_{0,k}D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})$*. We have that* $(\bar{M}(k) : k = 1, \ldots)$ *is a discrete martingale w.r.t.* $\mathcal{F}_k = (\mathbf{O}_1, \ldots, \mathbf{O}_k)$*:* $E_0(\bar{M}(K) \mid \mathcal{F}(k)) = \bar{M}(k)$ *for* $k \leq K$*. Let*

$$\Sigma_k^2 \equiv E_0 M_k^2 \equiv E_0 M_k M_k^\top,$$

and

$$\Sigma^2(K) \equiv \frac{1}{K}\sum_{k=1}^{K}\Sigma_k^2.$$

Define also

$$W^2(K) = \frac{1}{K}\sum_{k=1}^{K} E_0(M_k^2 \mid \mathcal{F}_{k-1}) = \frac{1}{K}\sum_{k=1}^{K} P_{0,k}M_k^2.$$

Define

$$\bar{R}(K) \equiv \frac{1}{K}\sum_{k=1}^{K} R(Q_{k-1}, G_{k-1}, Q_0, G_0, \mathbf{g}_{0,k}).$$

We have the following expansion for the online one-step estimator:

$$\psi_K - \psi_0 = \frac{\bar{M}(K)}{K} + \bar{R}(K).$$

Assumptions: *We make the following assumptions*

- *For some* $M < \infty$ $\max_k \mid D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k})(\mathbf{O}_k) \mid < M < \infty$ *with probability 1;*
- $\bar{R}(K) = o_P(1/\sqrt{K})$*;*
- $\liminf_{K\to\infty} \lambda\Sigma^2(K)\lambda > 0$ *for all* λ*, or that*

  $$\Sigma^2 = \lim_{k\to\infty} \Sigma(k)^2 \text{ exists and is a positive definite covariance matrix;}$$

- $W^2(K) - \Sigma^2(K) \to_{K\to\infty} 0$ *in probability, or, equivalently*

  $$\frac{1}{K}\sum_{k=1}^{K} P_{0,k} D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k}) - E_0 \frac{1}{K}\sum_{k=1}^{K} P_{0,k} D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k}) \to 0 \tag{12}$$

  in probability as $K \to \infty$*.*

Conclusion: *Then,*

$$\Sigma(K)^{-1}\frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0, I), \text{ as } K \to \infty,$$

and, if $\Sigma^2(K) \to \Sigma^2$*, as* $K \to \infty$*, for some positive definite covariance matrix* Σ^2*, then*

$$\frac{\bar{M}(K)}{\sqrt{K}} \Rightarrow_D N(0, \Sigma^2), \text{ as } K \to \infty.$$

This implies:

$$\sqrt{Km}(\psi_K - \psi_0) \Rightarrow_D N(0, \Sigma^2/m) \text{ as } K \to \infty.$$

If $P_{0,k}\{D_k^*(Q_{k-1}, G_{k-1}, \mathbf{g}_{0,k}) - D^*(Q_0, G_0, g_0)\}^2 \to 0$ *in probability, as* $k \to \infty$*, then* $\Sigma^2/m = P_0 D^*(Q_0, G_0, g_0)^2$ *is the efficiency bound for i.i.d. sampling from*

P_{θ_0,g_0}, which proves that ψ_K is asymptotically equivalent with an asymptotically efficient estimator under i.i.d. sampling from P_{θ_0,g_0}.
Consistent estimation of asymptotic variance: *Finally, consider the following estimator of $\Sigma^2(K)$:*

$$\hat{\Sigma}^2(K) = \frac{1}{K}\sum_{k=1}^{K}\{D_k^*(Q_{k-1}, G_{k-1}(\mathbf{O}_k) - \bar{D}_K\}^2,$$

where $\bar{D}_K = \frac{1}{K}\sum_{k=1}^{K} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k)$. We have $\hat{\Sigma}^2(K) - \Sigma^2(K) \to 0$ in probability as $K \to \infty$, and if Σ^2 exists, then we also have $\hat{\Sigma}^2(K) \to \Sigma^2$ in probability as $K \to \infty$.

The proof of Theorem 7 is completely identical to the proof of Theorem 1.

10.4 Online TMLE

As is clear from the previous section, our online TMLE can also be straightforwardly generalized to group sequential adaptive designs and the same theorems apply to this more general setting. So we can conclude that since our results in the previous sections only relied on a martingale structure of the estimators, group sequential adaptive designs are naturally captured by our set up and theorems without any extra work.

11 Discussion

This article concerned the generalization of targeted learning for i.i.d. data, modeled by large semi parametric models, to online targeted learning in order to scale up these methods to handle massive amount of possibly streaming data. Specifically, we proposed methods for online super-learning, online one-step and online targeted minimum loss-based estimation. In addition, we also generalized our online semi-parametric efficient estimators to group sequential data adaptive designs, as studied with regular TMLE in van der Laan (2008); Chambaz and van der Laan (2011a,b). This article provides a solid basis for developing online estimators for specific estimation problems, as demonstrated in our treatment specific mean example. The framework appears to be general enough to incorporate new and innovative ideas of how to construct online estimators that are maximally computationally friendly without giving up on the fundamental statistical properties such as asymptotic linearity, efficiency, and global finite sample robustness. Due to the martingale structure of our online estimators, their asymptotic does not rely on the usual Donsker class condition, allowing them to be asymptotically linear and efficient under weaker conditions than typical estimators.

In the future, the theoretical properties of our online super-learning approach will need to be established. Our online TMLE relies on evaluation of $\Psi(Q_k(\epsilon_k^*))$ at each step k and if Q_k has a dimension growing with sample size, then this estimator is more computer intensive than our online one-step estimator utilizing an online evaluation ψ_k of $\Psi()$ (thereby never evaluating the Ψ of the current estimator, but just updating the previous ψ_{k-1}). That is, it appears that insisting on an estimator to be a substitution estimator $\Psi(Q_k^*)$ one naturally slows down its computation due to having to evaluate $\Psi()$ at each step, and the number of calculations needed to evaluate $\Psi(Q_k^*)$ will depend on the dimension of Q_k^*. On the other hand, if the dimension of the initial online estimator Q_k grows with sample size n_k, then this extra computation required to evaluate $\Psi(Q_k^*)$ is of the same order as computing the initial online Q_k, so that the computational gain of the online one step estimator is not that meaningful.

We have suggested that the online targeted one-step estimator relying on targeted initial estimators $Q_k(\epsilon_k^0)$ for a well constructed online second order SGD estimator ϵ_k^0, combined with an online evaluation ψ_{k-1}^0 of $\Psi()$ might be asymptotically equivalent with $1/K\sum_{k=1}^{K}\psi_{k-1}^0$, so that the latter can be used instead, inheriting the robustness properties of a substitution estimator. Clearly, more theoretical work remains to be done to confirm this conjecture and to establish if our current proposed online targeted one step estimators do the job. In that manner, one essentially still obtains robust substitution estimators that are as fast to compute as the online one-step estimator.

An interesting future challenge will be the construction of an online estimator of a dynamic treatment specific mean EY_d for a dynamic treatment multiple time point intervention d based on a general longitudinal data structure, relying on the sequential regression representation, thereby creating an online efficient estimators analogue to the TMLE in (Gruber and van der Laan, 2012; Petersen et al., 2013), where the latter TMLEs are inspired by important double robust estimators established in earlier work of Bang and Robins (2005). Finally, we will need to implement corresponding online R packages analogue to the super-learner package and the longitudinal TMLE package (superlearner(), ltmle()).

Acknowledgement

This research was supported in part by NIH grant 2R01AI074345.

Appendix

Proof of Theorem 5.

Consider the online TMLE $\psi_K^* = \Psi(Q_K(\epsilon_K^*))$ as estimator of ψ_0, and the conclusion of Theorem 4 regarding $\epsilon_K^* - \epsilon_{0K}$. We have for an $\tilde{\epsilon}_{0K}$ between ϵ_K^* and ϵ_{0K}

$$\begin{aligned}
\Psi(Q_K(\epsilon_K^*)) - \Psi(Q_0) &= \Psi(Q_K(\epsilon_K^*)) - \Psi(Q_K(\epsilon_{0K})) \\
&= \frac{d}{d\tilde{\epsilon}_{0K}} \Psi(Q_K(\tilde{\epsilon}_{0K}))(\epsilon_K^* - \epsilon_{0K})) \\
&= \frac{d}{d\tilde{\epsilon}_{0K}} \Psi(Q_K(\epsilon_{0K}))c_{0K}^{-1} \frac{\bar{M}(K)}{K} + o_P(1/\sqrt{K}) \\
&= \frac{\bar{M}(K)}{K} + o_P(1/\sqrt{K}),
\end{aligned}$$

where we use that $\tilde{\epsilon}_{0K} - \epsilon_{0K}$ converges to zero as $K \rightarrow \infty$, so that, by assumption

$$\frac{d}{d\tilde{\epsilon}_{0K}} \Psi(Q_K(\tilde{\epsilon}_{0K}))c_{0K}^{-1} \rightarrow 1 \text{ in probabilty as } K \rightarrow \infty.$$

Proof of Theorem 4.

Let's now analyze this estimator of ϵ_{0K}. We have

$$\begin{aligned}
&\epsilon_K^* = \tfrac{1}{K} \textstyle\sum_{k=1}^K c_K^{-1} \Psi(Q_{k-1}) \\
&+ \tfrac{1}{K} \textstyle\sum_{k=1}^K \left\{ \tilde{f}_K^{-1} D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - \tilde{f}_K^{-1} P_{0k} D_k^*(Q_{k-1}, G_{k-1}) \right\} \\
&+ \tfrac{1}{K} \textstyle\sum_{k=1}^K \tilde{f}_K^{-1} P_{0k} D_k^*(Q_{k-1}, G_{k-1}) \\
&= \tfrac{1}{K} \textstyle\sum_{k=1}^K c_K^{-1} \Psi(Q_{k-1}) + c_K^{-1} \tfrac{M(K)}{K} \\
&+ \tfrac{1}{K} \textstyle\sum_{k=1}^K \tilde{f}_K^{-1} \{\Psi(Q_0) - \Psi(Q_{k-1}) + R_{0k}(Q_{k-1}, G_{k-1}, Q_0, G_0)\},
\end{aligned}$$

where

$$\bar{M}(K) = \sum_{k=1}^K \{D_k^*(Q_{k-1}, G_{k-1})(\mathbf{O}_k) - P_{0k} D_k^*(Q_{k-1}, G_{k-1})\},$$

and we used $\tilde{f}_K^{-1}a - \tilde{f}_K^{-1}b = c_K^{-1}(a-b)$ for any two numbers a, b. Let's now focus on the last term, which equals

$$\begin{aligned}
&\tfrac{1}{K}\textstyle\sum_{k=1}^K \left\{\epsilon_K^0 - c_K^{-1} f_K(\epsilon_K^0) + c_K^{-1}\psi_0 - c_K^{-1}\Psi(Q_{k-1}) + c_K^{-1} R_{0k}\right\} \\
&= \{\epsilon_K^0 - c_K^{-1} f_K(\epsilon_K^0) + c_K^{-1}(\psi_0)\} - \tfrac{1}{K}\textstyle\sum_{k=1}^K c_K^{-1}(\Psi(Q_{k-1}) + c_K^{-1}\tfrac{1}{K}\textstyle\sum_{k=1}^K R_{0k} \\
&= \tilde{f}_K^{-1}(\psi_0) - \tfrac{1}{K}\textstyle\sum_{k=1}^K c_K^{-1}\Psi(Q_{k-1}) + c_K^{-1}\bar{R}(K),
\end{aligned}$$

where $\bar{R}(K) \equiv \frac{1}{K}\sum_{k=1}^K R_{0k}(Q_{k-1}, G_{k-1}, Q_0, G_0)$. Plug this expression back into our expression above for ϵ_K^* to obtain:

$$\epsilon_K^* - \tilde{f}_K^{-1}(\psi_0) = c_K^{-1}\frac{\bar{M}(K)}{K} + c_K^{-1}\bar{R}(K).$$

Finally, we use that $\tilde{f}_K^{-1}(\psi_0) = f_K^{-1}(\psi_0) + R_0(\epsilon_K^0, \epsilon_{0K})$ and $f_K^{-1}(\psi_0) = \epsilon_{0K}$ to obtain:

$$\epsilon_K^* - \epsilon_{0K} = c_K^{-1}\frac{\bar{M}(K)}{k} + c_K^{-1}\bar{R}(K) + R_0(\epsilon_K^0, \epsilon_{0K}).$$

We assumed that $\bar{R}(K) = o_P(1/\sqrt{K})$ and $R_0(\epsilon_K^0, \epsilon_{0K}) = o_P(1/\sqrt{K})$. Under these assumptions, we have

$$\epsilon_K^* - \epsilon_{0K} = c_K^{-1}\frac{\bar{M}(K)}{K} + o_P(1/\sqrt{K}).$$

This completes the proof of Theorem 4.

References

J. Andersen, D. Faries, and R. Tamura. A randomized play-the-winner design for multiarm clinical trials. *Communication in Statistical Theory*, 23:309–323, 1994.

Z.D. Bai, F. Hu, and W.F. Rosenberger. Asymptotic properties of adaptive designs for clinical trials with delayed response. *Annals of Statistics*, 30(1): 122–139, 2002.

H. Bang and J.M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–972, 2005.

P.J. Bickel, C.A.J. Klaassen, Y. Ritov, and J. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Springer-Verlag, 1997.

Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

Léon Bottou. Stochastic gradient descent tricks. In *Neural Networks: Tricks of the Trade*, pages 421–436. Springer, 2012.

A. Chambaz and M.J. van der Laan. Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate, theoretical study. *Int J Biostat*, 7(1):1–32, 2011a. Working paper 258, www.bepress.com/ucbbiostat.

A. Chambaz and M.J. van der Laan. Targeting the optimal design in randomized clinical trials with binary outcomes and no covariate, simulation study. *Int J Biostat*, 7(1):33–, 2011b. Working paper 258,www.bepress.com/ucbbiostat.

A. Chambaz, W. Zheng, and M.J. van der Laan. Targeted covariate-adjusted response-adaptive lasso-based randomized controlled trials. Technical Report 323, Division of Biostatistics, University of California, Berkeley, 2014.

Y. Cheng and Y. Shen. Bayesian adaptive designs for clinical trials. *Biometrika*, 92(3):633–646, 2005.

John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *The Journal of Machine Learning Research*, 999999:2121–2159, 2011.

N. Flournoy and W.F. Rosenberger. *Adaptive Designs*. Hayward, Institute of Mathematical Statistics, 1995.

S. Gruber and M.J. van der Laan. Targeted minimum loss based estimator that outperforms a given estimator. *The International Journal of Biostatistics*, 8(1):Article 11, doi: 10.1515/1557–4679.1332, 2012.

P.W. Holland. Statistics and causal inference. *J Am Stat Assoc*, 81(396): 945–960, 1986.

F. Hu and W.F. Rosenberger. Analysis of time trends in adaptive designs withi appliation to a neurophysiology experiment. *Statistics in Medicine*, 19:2067–2075, 2000.

Noboru Murata. A statistical study of on-line learning. *Online Learning and Neural Networks. Cambridge University Press, Cambridge, UK*, 1998.

J. Neyman. On the application of probability theory to agricultural experiments. *Statistical Science*, 5:465–480, 1990.

J. Pearl. *Causality: models, reasoning, and inference.* Cambridge, New York, 2nd edition, 2009.

M. Petersen, J. Schwab, S. Gruber, N. Blaser, M. Schomaker, and M.J. van der Laan. Targeted minimum loss based estimation of marginal structural working models. *Journal of Causal Inference*, submitted, technical report http://biostats.bepress.com/ucbbiostat/paper312/, 2013.

E.C. Polley, Sherri Rose, and M.J. van der Laan. Super learning. In M.J. van der Laan and S. Rose, editors, *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer, New York Dordrecht Heidelberg London, 2012.

Boris T Polyak and Anatoli B Juditsky. Acceleration of stochastic approximation by averaging. *SIAM Journal on Control and Optimization*, 30(4): 838–855, 1992.

J.M. Robins. Addendum to: "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect" [Math. Modelling **7** (1986), no. 9-12, 1393–1512; MR 87m:92078]. *Comput. Math. Appl.*, 14(9-12):923–945, 1987a. ISSN 0097-4943.

J.M. Robins. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. *J Chron Dis (40, Supplement)*, 2:139s–161s, 1987b.

J.M. Robins and A. Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS Epidemiology*, Methodological issues. Bikhäuser, 1992.

S. Rose and M.J. van der Laan. *Targeted Learning: Causal Inference for Observational and Experimental Data.* Springer, New York, 2011.

W.F. Rosenberger. New directions in adaptive designs. *Statistical Science*, 11: 137–149, 1996.

W.F. Rosenberger and S.E. Grill. A sequential design for psychophysical experiments: An application to estimating timing of sensory events. *Statistics in Medicine*, 16:2245–2260, 1997.

W.F. Rosenberger and T.N. Shiram. Estimation for an adaptive allocation design. *Journal of Statistical Planning and Inference*, 59:309–319, 1997.

W.F. Rosenberger, N. Flournoy, and S.D. Durham. Asymptotic normality of maximum likelihood estimators from multiparameter response driven designs. *Journal of Statistical Planning and Inference*, pages 69–76, 1997.

D.B. Rubin. Estimating causal effects of treatments in randomized and non-randomized studies. *J. Educ Psychol*, 64:688–701, 1974.

D.B. Rubin. *Matched Sampling for Causal Effects.* Cambridge University Press, Cambridge, MA, 2006.

P.K. Sen and J.M. Singer. *Large Sample Methods in Statistics: An Introduction with Applications*. Chapman & Hall, New York, 1993.

R.N. Tamura, D.E. Faries, J.S. Andersen, and J.H. Heiligenstein. A case study of an adaptive clinical trial in the treatment of out-patients with depressive disorder. *Journal of the American Statistical Association*, 89:768–776, 1994.

M.J. van der Laan. Estimation based on case-control designs with known prevalance probability. *The International Journal of Biostatistics*, page http://www.bepress.com/ijb/vol4/iss1/17/, 2008.

M.J. van der Laan and S. Dudoit. Unified cross-validation methodology for selection among estimators and a general cross-validated adaptive epsilon-net estimator: Finite sample oracle inequalities and examples. Technical report, Division of Biostatistics, University of California, Berkeley, November 2003.

M.J. van der Laan and J.M. Robins. *Unified methods for censored longitudinal data and causality*. Springer, New York, 2003.

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

M.J. van der Laan, S. Dudoit, and A.W. van der Vaart. The cross-validated adaptive epsilon-net estimator. *Statistics and Decisions*, 24(3):373–395, 2006.

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

A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

A.W. van der Vaart, S. Dudoit, and M.J. van der Laan. Oracle inequalities for multi-fold cross-validation. *Statistics and Decisions*, 24(3):351–371, 2006.

L.J. Wei. The generalized polya's urn design for sequential medical trials. *Annals of Statistics*, 7:291–296, 1979.

L.J. Wei and S. Durham. The randomize play-the-winner rule in medical trials. *Journal of the American Statistical Association*, 73:840–843, 1978.

L.J. Wei, R.T. Smythe, D.Y. Lin, and T.S. Park. Statistical inference with data-dependent treatment allocation rules. *Journal of the American Statistical Association*, 85:156–162, 1990.

Wei Xu. Towards optimal one pass large scale learning with averaged stochastic gradient descent. *CoRR*, abs/1107.2490, 2011. URL `http://arxiv.org/abs/1107.2490`.

Matthew D Zeiler. Adadelta: An adaptive learning rate method. *arXiv preprint arXiv:1212.5701*, 2012.

M. Zelen. Play the winner rule and the controlled clinical trial. *Journal of the American Statistical Association*, 64:131–146, 1969.