The state learner – a super learner for right-censored data

Anders Munch & Thomas A. Gerds

May 21, 2025

Abstract

In survival analysis, prediction models are needed as stand-alone tools and in applications of causal inference to estimate nuisance parameters. The super learner is a machine learning algorithm which combines a library of prediction models into a meta learner based on cross-validated loss. In right-censored data, the choice of the loss function and the estimation of the expected loss need careful consideration. We introduce the state learner, a new super learner for survival analysis, which simultaneously evaluates libraries of prediction models for the event of interest and the censoring distribution. The state learner places no restrictions on the algorithms that can be included in the library, works in the presence of competing risks, and does not require a single pre-specified estimator of the conditional censoring distribution. We establish an oracle inequality for the state learner and investigate its performance through numerical experiments. We illustrate the application of the state learner with prostate cancer data.

Keywords: *Competing risks, cross-validation, loss based estimation, right-censored data, super learner*

1 Introduction

[More motivation]

A super learner is a machine learning algorithm that combines a finite set of learners into a meta learner by estimating prediction performance in hold-out samples using a pre-specified loss function [van der Laan et al., 2007]. When the aim is to make a prediction model, super learners combine strong learners, such as Cox regression models and random survival forests [Gerds and Kattan, 2021, Section 8.4]. While the general idea of combining strong learners based on cross-validation stems from earlier work [Wolpert, 1992, Breiman, 1996], the name super learner is justified by an oracle inequality [van der Laan and Dudoit, 2003, van der Vaart et al., 2006].

We define the state learner, a new super learner for right-censored data, which simultaneously estimates the expected loss of learners of the event time distribution and the censoring distribution. Compared to the super learners of Polley and van der Laan [2011] and Golmakani and Polley [2020], the state learner proposed here poses no restrictions on the type of learners that can be included in the library. Compared to super learners that rely on inverse probability of censoring weighting, censoring unbiased transformations, or pseudo-values, the state learner does not require a pre-specified censoring distribution. Compared to the super learners by Han et al. [2021] and Westling et al. [2021], the state learner has theoretical guarantees which these methods do not have.

The loss function underlying the state learner operates on coarsened data. In the right-censored survival setting the coarsened data consist of the minimum and the order of the censoring time and the event time as well as the baseline covariates. The state learner works by simultaneously assessing how well learners of the event time distribution and the censoring distribution predict the states of the coarsened data. To analyse the theoretical properties of the state learner we focus on the discrete super learner which combines the library of learners by picking the one that minimises the cross-validated loss [van der Laan et al., 2007]. In the presence of competing risks, our algorithm uses separate libraries of learners for the cumulative hazard functions for each of the competing risks and for the censoring distribution. We show that the oracle selector of the state learner is consistent if all libraries contain a consistent learner and prove a finite sample oracle inequality.

The state learner algorithm can output a medical risk prediction model [Gerds and Kattan, 2021] which predicts the probability of an event based on covariates in the presence of competing risks. Another important application, which motivated our contribution, is in targeted learning where conditional event probabilities occur as high-dimensional nuisance parameters which need to be estimated at a certain rate [van der Laan and Rose, 2011, Rytgaard et al., 2021, Rytgaard and van der Laan, 2022]. The state learner is particularly well-suited for this application as it simultaneously provides estimates for both the outcome and the censoring, both of which are needed for targeted learning. For the asymptotic bias term of a targeted estimator, which uses the state learner to estimate nuisance parameters, we show that a second order product structure holds. We illustrate both applications of the state learner with prostate cancer data [Kattan et al., 2000].

There are many existing super learners for right-censored data. We briefly review them here. Machine learning based on right-censored data commonly uses the partial log-likelihood as a loss function [e.g., Li et al., 2016, Yao et al., 2017, Lee et al., 2018, Katzman et al., 2018, Gensheimer and Narasimhan, 2019, Lee et al., 2021, Kvamme and Borgan, 2021]. This loss function is also suggested for super learning with right-censored data by Polley and van der Laan [2011], where data is assumed to be observed in discrete time. However, the partial log-likelihood loss does not work well with data splitting (cross-validation) in continuous time. The reason is that the partial log-likelihood loss assigns an infinite value for any learner that predicts piece-wise constant cumulative hazard functions when there are event times in the test set which do not occur in the learning set. This problem occurs with prominent survival learners including the Kaplan-Meier estimator, the random survival forest, and the semi-parametric Cox regression model and these learners cannot be included in the library of the super learner proposed by Polley and van der Laan [2011]. When a proportional hazards model is assumed, the baseline hazard function can be profiled out of the likelihood [Cox, 1972]. The cross-validated partial log-likelihood loss [Verweij and van Houwelingen, 1993] has therefore been suggested as a loss function for super learning by Golmakani and Polley [2020]. This choice of loss function restricts the library of learners to include only Cox proportional hazards models, and hence excludes many learners such as, e.g., random survival forests, additive hazards models, and accelerated failure time models. Other proposals for super learning with right-censored data use an inverse probability of censoring weighted (IPCW) loss function [Graf et al., 1999, van der Laan and Dudoit, 2003, Molinaro et al., 2004, Keles et al., 2004, Hothorn et al., 2006, Gerds and Schumacher, 2006, Gonzalez Ginestet et al., 2021], censoring unbiased transformations [Fan and Gijbels, 1996, Steingrimsson et al., 2019], or pseudo-values [Andersen et al., 2003, Mogensen and Gerds, 2013, Sachs et al., 2019]. All these methods rely on an estimator of the censoring distribution, and their drawback is that this estimator has to be pre-specified. An approach which avoids a pre-specified censoring model was proposed independently by Han et al. [2021] and Westling et al. [2021]. In both articles, the authors suggest to iterate between learning of the outcome model and learning of the censoring model using IPCW loss functions. However, it seems that this procedure is in general not guaranteed to converge to the true data-generating mechanism [Munch, 2024, Appendix A.4].

We introduce our notation and framework in Section 2. In Section 3 we define general super learning for right-censored data. Section 4 introduces the state learner, and Section 6 provides theoretical guarantees. We report results of our numerical experiments in Section 7 and analyse a prostate cancer data set in Section 8. Section 9 contains a discussion of the merits and limitations of our proposal. Proofs are given in Appendix A. An implementation of the state learner is available at `https://github.com/amnudn/statelearner` along with code for reproducing our numerical experiments.

2 Notation and framework

In a competing risk framework [Andersen et al., 2012], let T be a time to event variable, $D \in \{1, 2\}$ the cause of the event, and $X \in \mathcal{X}$ a vector of baseline covariates taking values in a bounded subset $\mathcal{X} \subset \mathbb{R}^p$, $p \in \mathbb{N}$. Let $\tau < \infty$ be the prediction horizon. We use $\mathcal{Q}$ to denote the collection of all probability measures on $[0, \tau] \times \{1, 2\} \times \mathcal{X}$ such that $(T, D, X) \sim Q$ for some unknown $Q \in \mathcal{Q}$. For $j \in \{1, 2\}$, the cause-specific conditional cumulative hazard functions are defined by $\Lambda_j \colon [0, \tau] \times \mathcal{X} \to \mathbb{R}_+$ such that

$$\Lambda_j(t \mid x) = \int_0^t \frac{Q(T \in \mathrm{d}s, D = j \mid X = x)}{Q(T \geq s \mid X = x)}.$$

For ease of presentation we assume throughout that the map $t \mapsto \Lambda_j(t \mid x)$ is continuous for all x and j. This is not a limitation: All arguments carry over directly to the general case. We denote by S the conditional event-free survival function:

$$S(t \mid x) = \exp\left\{-\Lambda_1(t \mid x) - \Lambda_2(t \mid x)\right\}. \tag{1}$$

Let $\mathcal{M}_\tau$ denote the space of all conditional cumulative hazard functions on $[0, \tau] \times \mathcal{X}$. Any distribution $Q \in \mathcal{Q}$ can be characterised by

$$\begin{aligned} Q(\mathrm{d}t, j, \mathrm{d}x) = &\{S(t- \mid x)\Lambda_1(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=1\}} \\ &\{S(t- \mid x)\Lambda_2(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=2\}}, \end{aligned}$$

where $\Lambda_j \in \mathcal{M}_\tau$ for $j = 1, 2$ and H is the marginal distribution of the covariates.

We consider the right-censored setting in which we observe the coarsened data $O = (\tilde{T}, \tilde{D}, X)$, where $\tilde{T} = \min(T, C)$ for a right-censoring time C, $\Delta = \mathbb{1}\{T \leq C\}$, and $\tilde{D} = \Delta D$. Let $\mathcal{P}$ denote a set of probability measures on the sample space $\mathcal{O} = [0, \tau] \times \{0, 1, 2\} \times \mathcal{X}$ such that $O \sim P$ for some unknown $P \in \mathcal{P}$. We assume that the event times and the censoring times are conditionally independent given covariates, $T \perp\!\!\!\perp C \mid X$. This implies that any distribution $P \in \mathcal{P}$ is characterised by a distribution $Q \in \mathcal{Q}$ and a conditional cumulative hazard function for C given X [c.f., Begun et al., 1983, Gill et al., 1997]. We use $\Gamma \in \mathcal{M}_\tau$ to denote the conditional cumulative hazard function for censoring. For ease of presentation we now also assume that $\Gamma(\cdot \mid x)$ is continuous for all x. We let $(t, x) \mapsto G(t \mid x) = \exp\{-\Gamma(t \mid x)\}$ denote the survival function of the conditional censoring distribution. The distribution P is characterised by

$$
\begin{aligned}
P(\mathrm{d}t, j, \mathrm{d}x) = &\{G(t- \mid x)S(t- \mid x)\Lambda_1(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=1\}} \\
&\{G(t- \mid x)S(t- \mid x)\Lambda_2(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=2\}} \\
&\{G(t- \mid x)S(t- \mid x)\Gamma(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=0\}} \\
= &\{G(t- \mid x)Q(\mathrm{d}t, j, \mathrm{d}x)\}^{\mathbb{1}\{j\neq 0\}} \\
&\{G(t- \mid x)S(t- \mid x)\Gamma(\mathrm{d}t \mid x)H(\mathrm{d}x)\}^{\mathbb{1}\{j=0\}}.
\end{aligned} \tag{2}
$$

Hence, we may write $\mathcal{P} = \{P_{Q,\Gamma} : Q \in \mathcal{Q}, \Gamma \in \mathcal{G}\}$ for some $\mathcal{G} \subset \mathcal{M}_\tau$. We also have H-almost everywhere

$$
P(\tilde{T} > t \mid X = x) = S(t \mid x)G(t \mid x) = \exp\{-\Lambda_1(t \mid x) - \Lambda_2(t \mid x) - \Gamma(t \mid x)\}.
$$

We further assume that there exists $\kappa < \infty$ such that $\Lambda_j(\tau- \mid x) < \kappa$, for $j \in \{1, 2\}$, and $\Gamma(\tau- \mid x) < \kappa$ for almost all $x \in \mathcal{X}$. Note that this implies that $G(\tau- \mid x)$ is bounded away from zero for almost all $x \in \mathcal{X}$. Under these assumptions, the conditional cumulative hazard functions Λ_j and Γ can be identified from P by

$$
\Lambda_j(t \mid x) = \int_0^t \frac{P(\tilde{T} \in \mathrm{d}s, \tilde{D} = j \mid X = x)}{P(\tilde{T} \geq s \mid X = x)}, \tag{3}
$$

$$
\Gamma(t \mid x) = \int_0^t \frac{P(\tilde{T} \in \mathrm{d}s, \tilde{D} = 0 \mid X = x)}{P(\tilde{T} \geq s \mid X = x)}. \tag{4}
$$

Thus, we can consider Λ_j and Γ as operators which map from $\mathcal{P}$ to $\mathcal{M}_\tau$.

3 Super learning

In survival analysis, a super learner can be used to estimate a parameter $\theta \in \Theta$ which can be identified from the observed data distribution $P \in \mathcal{P}$. A super learner typically estimates a function, and thus the parameter space Θ is most often a class of functions. For example, θ could be a conditional survival function or a cumulative hazard function. In Section 4 we consider a specific choice of parameter as target for our super learner, but in this section we describe super learning for a general parameter θ with values in some parameter space Θ.

As input to a super learner we need a data set $\mathcal{D}_n = \{O_i\}_{i=1}^n$ of i.i.d. observations from $P \in \mathcal{P}$ and a collection of candidate learners $\mathcal{A}$. Each learner $a \in \mathcal{A}$ is a map $a \colon \mathcal{O}^n \to \Theta$ which takes a data set as input and returns an estimate $a(\mathcal{D}_n) \in \Theta$ of θ. In what follows, we use the short-hand notation $P[f] = \int f(o)P(\mathrm{d}o)$. A super learner evaluates the performance of $a \in \mathcal{A}$ with a loss function $L \colon \Theta \times \mathcal{O} \to \mathbb{R}_+$ and estimates the expected loss $P[L(a(\mathcal{D}_n), \cdot)]$ using cross-validation. Specifically, the expected loss of $a \in \mathcal{A}$ is estimated by splitting the data set $\mathcal{D}_n$ into K disjoint approximately equally sized subsets $\mathcal{D}_n^1, \mathcal{D}_n^2, \dots, \mathcal{D}_n^K$ and then calculating the cross-validated loss

$$
\hat{R}_n(a; L) = \frac{1}{K}\sum_{k=1}^K \frac{1}{|\mathcal{D}_n^k|}\sum_{O_i \in \mathcal{D}_n^k} L\big(a(\mathcal{D}_n^{-k}), O_i\big), \quad \text{with} \quad \mathcal{D}_n^{-k} = \mathcal{D}_n \setminus \mathcal{D}_n^k.
$$

The subset $\mathcal{D}_n^{-k}$ is referred to as the k'th training sample, while $\mathcal{D}_n^k$ is referred to as the k'th test or hold-out sample. The discrete super learner is defined as

$$
\hat{a}_n = \operatorname*{argmin}_{a \in \mathcal{A}} \hat{R}_n(a; L).
$$

The oracle learner is defined as the learner that minimises the expected loss under the data-generating distribution P, i.e.,

$$\tilde{a}_n = \operatorname*{argmin}_{a \in \mathcal{A}} \tilde{R}_n(a; L), \quad \text{with} \quad \tilde{R}_n(a; L) = \frac{1}{K} \sum_{k=1}^{K} P\big[L\big(a(\mathcal{D}_n^{-k}), \cdot\big)\big].$$

Note that both the discrete super learner and the oracle learner depend on the library of learners and on the number of folds K, and that the oracle learner is a function of the data and the unknown data-generating distribution. However, these dependencies are suppressed in the notation.

4 The state learner

The main idea of the state learner is to jointly use learners for Λ_1, Λ_2, and Γ, and the relations in equation (2), to learn a feature of the observed data distribution P. The discrete state learner ranks a tuple of learners for the tuple of the cumulative hazard functions $(\Lambda_1, \Lambda_2, \Gamma)$ based on how well they jointly model the observed data. To formally introduce the state learner, we define the multi-state process

$$\eta(t) = \mathbb{1}\{\tilde{T} \leq t, \tilde{D} = 1\} + 2\,\mathbb{1}\{\tilde{T} \leq t, \tilde{D} = 2\} - \mathbb{1}\{\tilde{T} \leq t, \tilde{D} = 0\}, \quad \text{for} \quad t \in [0, \tau].$$

At time t, we observe that each individual is in one of four mutually exclusive states: 0, 1, 2, or -1. The conditional distribution of the process $\eta(t)$ given baseline covariates X is determined by the function

$$F(t, k, x) = P(\eta(t) = k \mid X = x), \tag{5}$$

for all $t \in [0, \tau]$, $k \in \{-1, 0, 1, 2\}$, and $x \in \mathcal{X}$. The function F describes the conditional state occupation probabilities of the multi-state process η. We construct a super learner for F. The target of this super learner is the function-valued parameter $\theta(P) = F$ which is identified through equation (5). Under conditional independent censoring each tuple $(\Lambda_1, \Lambda_2, \Gamma, H)$ characterises a distribution $P \in \mathcal{P}$, c.f. equation (2), which in turn determines (F, H). Hence, a learner for F can be constructed from learners for Λ_1, Λ_2, and Γ as follows:

$$\begin{aligned}
F(t, 0, x) &= P(\tilde{T} > t \mid X = x) = \exp\{-\Lambda_1(t \mid x) - \Lambda_2(t \mid x) - \Gamma(t \mid x)\}, \\
F(t, 1, x) &= P(\tilde{T} \leq t, \tilde{D} = 1 \mid X = x) = \int_0^t F(s-, 0, x)\Lambda_1(\mathrm{d}s \mid x), \\
F(t, 2, x) &= P(\tilde{T} \leq t, \tilde{D} = 2 \mid X = x) = \int_0^t F(s-, 0, x)\Lambda_2(\mathrm{d}s \mid x), \\
F(t, -1, x) &= P(\tilde{T} \leq t, \tilde{D} = 0 \mid X = x) = \int_0^t F(s-, 0, x)\Gamma(\mathrm{d}s \mid x).
\end{aligned} \tag{6}$$

The state learner requires three libraries of learners: $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{B}$, where $\mathcal{A}_1$ and $\mathcal{A}_2$ contain learners for the conditional cause-specific cumulative hazard functions Λ_1 and Λ_2, respectively, and $\mathcal{B}$ contains learners for the conditional cumulative hazard function of the censoring distribution. Based on the Cartesian product of libraries of learners for $(\Lambda_1, \Lambda_2, \Gamma)$ we construct a library $\mathcal{F}$ of learners for F:

$$\mathcal{F}(\mathcal{A}_1, \mathcal{A}_2, \mathcal{B}) = \{\varphi_{a_1, a_2, b} : a_1 \in \mathcal{A}_1, a_2 \in \mathcal{A}_2, b \in \mathcal{B}\},$$

where in correspondence with the relations in equation (6),

$$\begin{aligned}
\varphi_{a_1,a_2,b}(\mathcal{D}_n)(t,0,x) &= \exp\left\{-a_1(\mathcal{D}_n)(s \mid x) - a_2(\mathcal{D}_n)(s \mid x) - b(\mathcal{D}_n)(s \mid x)\right\},\\
\varphi_{a_1,a_2,b}(\mathcal{D}_n)(t,1,x) &= \int_0^t \varphi_{a_1,a_2,b}(\mathcal{D}_n)(s-,0,x)a_1(\mathcal{D}_n)(\mathrm{d}s \mid x),\\
\varphi_{a_1,a_2,b}(\mathcal{D}_n)(t,2,x) &= \int_0^t \varphi_{a_1,a_2,b}(\mathcal{D}_n)(s-,0,x)a_2(\mathcal{D}_n)(\mathrm{d}s \mid x),\\
\varphi_{a_1,a_2,b}(\mathcal{D}_n)(t,-1,x) &= \int_0^t \varphi_{a_1,a_2,b}(\mathcal{D}_n)(s-,0,x)b(\mathcal{D}_n)(\mathrm{d}s \mid x).
\end{aligned}$$

To evaluate how well a function F predicts the observed multi-state process we use the integrated Brier score $\bar{B}_\tau(F,O) = \int_0^\tau B_t(F,O)\,\mathrm{d}t$, where B_t is the Brier score [Brier et al., 1950] at time $t \in [0,\tau]$,

$$B_t(F,O) = \sum_{j=-1}^{2} \left(F(t,j,X) - \mathbb{1}\{\eta(t) = j\}\right)^2.$$

Based on a split of a data set $\mathcal{D}_n$ into K disjoint approximately equally sized subsets (c.f., Section 3), each learner $\varphi_{a_1,a_2,b}$ in the library $\mathcal{F}(\mathcal{A}_1,\mathcal{A}_2,\mathcal{B})$ is evaluated using the cross-validated loss,

$$\hat{R}_n(\varphi_{a_1,a_2,b};\bar{B}_\tau) = \frac{1}{K}\sum_{k=1}^{K}\frac{1}{|\mathcal{D}_n^k|}\sum_{O_i\in\mathcal{D}_n^k}\bar{B}_\tau\left(\varphi_{a_1,a_2,b}(\mathcal{D}_n^{-k}),O_i\right),$$

and the discrete state learner is given by

$$\hat{\varphi}_n = \underset{(a_1,a_2,b)\in\mathcal{A}_1\times\mathcal{A}_2\times\mathcal{B}}{\operatorname{argmin}} \hat{R}_n(\varphi_{a_1,a_2,b};\bar{B}_\tau).$$

5 Use cases for the state learner

The state learner estimates the parameter F which is a feature of the observed right-censored data distribution P. In particular, F depends on the censoring distribution and is typically not of direct interest in itself. In the following, we describe how an estimate of F, as provided by the state learner, can be used to estimate more relevant parameters of interest, such as survival probabilities.

5.1 Survival and risk predictions

We here provide explicit formulas for obtaining event-free survival probabilities and risk predictions given the state-occupation function F defined in equation (5). These formula holds under the assumption of conditional independent censoring and positivity, as introduced in Section 2.

By equations (3) and (4) and the definition of F, we have

$$\Lambda_j(t,x) = \int_0^t \frac{F(\mathrm{d}s,j,x)}{F(s-,0,x)}, \quad j \in \{1,2\}. \tag{7}$$

Equation (1) provides a formula for obtaining the event-free survival probabilities from the cause-specific cumulative hazard functions. Similarly, cause-specific risk predictions can be obtained from Λ_1 and Λ_2 using the formula [e.g., Benichou and Gail, 1990, Ozenne et al., 2017],

$$Q(T \leq t, D = j \mid X = x) = \int_0^t \exp\{-\Lambda_1(u \mid x) - \Lambda_2(u \mid x)\} \Lambda_j(\mathrm{d}u \mid x), \quad j \in \{1, 2\}. \tag{8}$$

Hence, given the state learner's estimate of F we can use equation (7) to obtain estimates of the cause-specific cumulative hazard functions Λ_j, which can in turn be used to obtain estimates of the event-free survival probabilites and cause-specific risks through equations (1) and (8).

In Section 4 we suggested to implement the state learner by building a library using learners of the cause-specific cumulative hazard functions, Λ_j, and the cumulative hazard function for censoring, Γ. With this implementation we can directly input the highest ranked tuple of cause-specific hazard functions (Λ_1, Λ_2) provided by the state learner as input to equations (1) and (8).

5.2 Targeted and debiased machine learning

... TL + SL = strong... The state learner is particularly well suited for targeted and debiased machine learning as it immediately provides estimates of the nuisance parameters pertaining to the outcome and the censoring models. While it was our original motivation for developing the state learning, we leave the study of the state learner in the context of targeted and debiased machine learning for a future paper.

6 Theoretical results for the state learner

In this section we establish theoretical guarantees for the state learner. We show that the state learner is consistent if its library contains a consistent learner, and we establish a finite sample inequality for the excess risk of the state learner compared to the oracle.

6.1 Consistency

Proposition 6.1 can be derived from the fact that the integrated Brier score (also called the continuous ranked probability score) is a strictly proper scoring rule [Gneiting and Raftery, 2007]. This implies that if we minimise the average loss of the integrated Brier score, we recover the parameters of the data-generating distribution. Specifically, this implies that the oracle of a state learner is consistent if the library of learners contains at least one learner that is consistent for estimation of F. Recall that the function F implicitly depends on the data-generating probability measure $P \in \mathcal{P}$ but that this was suppressed in the notation. We now make this dependence explicit by writing F_0 for the function which is obtained by substituting a specific $P_0 \in \mathcal{P}$ for P in equation (6). In the following we let $\mathcal{H}_{\mathcal{P}} = \{F_P : P \in \mathcal{P}\}$ where F_P is defined as in equation (5) using the measure $P \in \mathcal{P}$.

Proposition 6.1. *If $P_0 \in \mathcal{P}$ then*

$$F_0 = \operatorname*{argmin}_{F \in \mathcal{H}_{\mathcal{P}}} P_0[\bar{B}_\tau(F, \cdot)],$$

H-almost surely for any $j \in \{-1, 0, 1, 2\}$ and almost any $t \in [0, \tau]$.

Proof. See Appendix A.1. □

6.2 Oracle inequalities

We establish a finite sample oracle result for the state learner. Our Corollary 6.2 is in essence a special case of Theorem 2.3 in [van der Vaart et al., 2006]. We assume that we split the data into equally sized folds, and for simplicity of presentation we take n to be such that $|\mathcal{D}_n^{-k}| = n/K$ with K fixed. We will allow the number of learners to grow with n and write $\mathcal{F}_n = \mathcal{F}(\mathcal{A}_{1,n}, \mathcal{A}_{2,n}, \mathcal{B}_n)$ as short-hand notation and to emphasise the dependence on n. In the following we let the space $\mathcal{H}_\mathcal{P}$ be equipped with the norm $\|\cdot\|_{P_0}$ defined as

$$\|F\|_{P_0} = \left\{ \sum_{j=-1}^{2} \int_\mathcal{X} \int_0^\tau F(t,j,x)^2 \,\mathrm{d}t H_0(\mathrm{d}x) \right\}^{1/2}. \tag{9}$$

Corollary 6.2. *For all $P_0 \in \mathcal{P}$, $n \in \mathbb{N}$, $k \in \{1, \dots, K\}$, and $\delta > 0$,*

$$\begin{aligned}\mathbb{E}_{P_0}\left[\|\hat{\varphi}_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] \leq{}& (1+2\delta)\,\mathbb{E}_{P_0}\left[\|\tilde{\varphi}_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] \\ &+ (1+\delta)16K\tau\left(13 + \frac{12}{\delta}\right)\frac{\log(1+|\mathcal{F}_n|)}{n}.\end{aligned}$$

Proof. See Appendix A.2. □

Corollary 6.2 has the following asymptotic consequences.

Corollary 6.3. *Assume that $|\mathcal{F}_n| = O(n^q)$, for some $q \in \mathbb{N}$ and that there exists a sequence $\varphi_n \in \mathcal{F}_n$, $n \in \mathbb{N}$, such that $\mathbb{E}_{P_0}\left[\|\varphi_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] = O(n^{-\alpha})$, for some $\alpha \leq 1$.*

(a) If $\alpha = 1$ then $\mathbb{E}_{P_0}\left[\|\hat{\varphi}_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] = O(\log(n)n^{-1})$.

(b) If $\alpha < 1$ then $\mathbb{E}_{P_0}\left[\|\hat{\varphi}_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] = O(n^{-\alpha})$.

Proof. See Appendix A.2. □

7 Numerical experiments

In this section we report results from a simulation study where we consider estimation of the conditional survival function. In the first part, we compare the state learner to two IPCW based discrete super learners that use either the Kaplan-Meier estimator or a Cox model to estimate the censoring probability [Gonzalez Ginestet et al., 2021]. In the second part we compare the state learner to the super learner proposed by Westling et al. [2021].

In both parts we use the same data-generating mechanism. We generate data according to a distribution motivated from a real data set in which censoring depends on the baseline covariates. We simulate data based on the prostate cancer study of Kattan et al. [2000]. The outcome of interest is the time to tumor recurrence, and five baseline covariates are used to predict outcome: prostate-specific antigen (PSA, ng/mL), Gleason score sum (GSS, values between 6 and 10), radiation dose (RD), hormone therapy (HT, yes/no) and clinical stage (CS, six values). The study was designed such that a patient's radiation dose depended on when the patient entered the study [Gerds et al., 2013]. This in turn implies that the time of censoring depends on the radiation dose. The data were re-analysed in [Gerds et al., 2013] where a sensitivity analysis was conducted based on simulated data. Here we use the same simulation setup, where event and censoring times are generated according to parametric Cox-Weibull models estimated from the original data, and the covariates are generated according to either marginal Gaussian normal or binomial distributions estimated from the original data [c.f., Gerds et al., 2013, Section 4.6]. We refer to this simulation setting as 'dependent censoring'. We also considered a simulation setting where data were generated in the same way, except that censoring was generated completely independently. We refer to this simulation setting as 'independent censoring'.

For all super learners we use a library consisting of three learners: The Kaplan-Meier estimator [Kaplan and Meier, 1958, Gerds, 2019], a Cox model with main effects [Cox, 1972, Therneau, 2022], and a random survival forest [Ishwaran et al., 2008, Ishwaran and Kogalur, 2023]. We use the same library to learn the outcome distribution and the censoring distribution. Note that the three learners in our library of learners can be used to learn the cumulative hazard functions of the outcome and the censoring distribution. The latter works by training the learner on the data set $\mathcal{D}_n^c$, where $\mathcal{D}_n^c = \{O_i^c\}_{i=1}^n$ with $O_i^c = (\tilde{T}_i, 1 - \Delta, X_i)$. When we say that we use a learner for the cumulative hazard function of the outcome to learn the cumulative hazard function of the censoring time, we mean that the learner is trained on $\mathcal{D}_n^c$.

We compare the state learner to two IPCW based super learners: The first super learner, called IPCW(Cox), uses a Cox model with main effects to estimate the censoring probabilities, while the second super learner, called IPCW(KM), uses the Kaplan-Meier estimator to estimate the censoring probabilities. The Cox model for the censoring distribution is correctly specified in both simulation settings while the Kaplan Meier estimator only estimates the censoring model correctly in the simulation setting where censoring is independent. Both IPCW super learners are fitted using the R-package `riskRegression` [Gerds et al., 2023]. The IPCW super learners use the integrated Brier score up to a fixed time horizon (36 months). The marginal risk of the event before this time horizon is $\approx 24.6\%$. Under the 'dependent censoring' setting the marginal censoring probability before the time horizon is $\approx 61.9\%$. Under the 'independent censoring' setting the marginal censoring probability before this time horizon is $\approx 38.7\%$.

Each super learner provides a learner for the cumulative hazard function for the outcome of interest. From the cumulative hazard function a risk prediction model can be obtained (c.f., equation (1) with $\Lambda_2 = 0$). We measure the performance of each super learner by calculating the index of prediction accuracy (IPA) [Kattan and Gerds, 2018] at a fixed time horizon (36 months) for the risk prediction model provided by the super learner. The IPA is 1 minus the ratio between the model's Brier score and the null model's Brier score, where the null model is the model that does not use any covariate information. The IPA is approximated using a large ($n = 20,000$) independent data set of uncensored data. As a benchmark we calculate the performance of the risk prediction model chosen by the oracle selector, which uses the large data set of uncensored event times to select the learner with the highest IPA.

The results are shown in Figure 1. We see that in the scenario where censoring depends on the covariates, using the Kaplan-Meier estimator to estimate the censoring probabilities provides a risk prediction model with an IPA that is lower than the risk prediction model provided by the state learner. The performance of the risk prediction model selected by the state learner is similar to the risk prediction model selected by the IPCW(Cox) super learner which a priori uses a correctly specified model for the censoring distribution. Both these risk prediction models are close to the performance of the oracle, except for small sample sizes.

We next compare the state learner to the super learner `survSL` [Westling et al., 2021]. This is another super learner which like the state learner works without a pre-specified censoring model. Note that both the state learner and `survSL` provide a prediction model for the

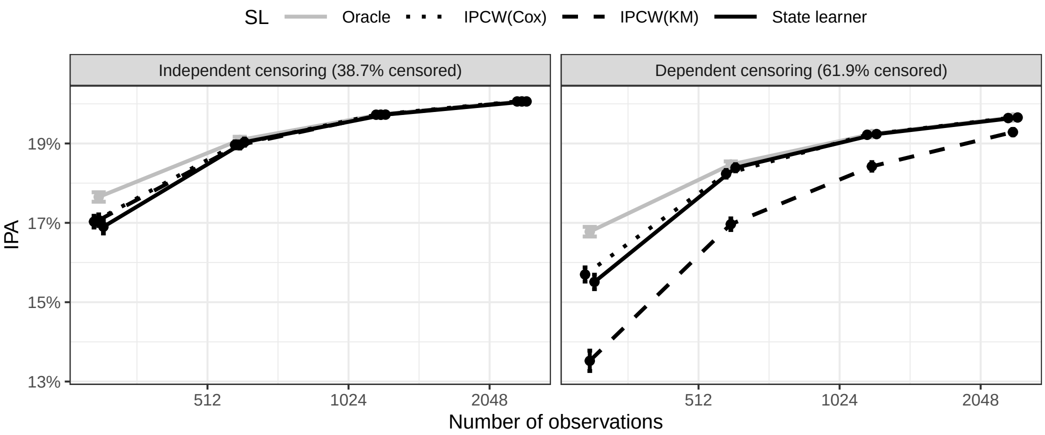

Figure 1: For the risk prediction models provided by each of the super learners, the IPA is plotted against sample size. The results are averages across 1000 simulated data sets and the error bars are used to quantify the Monte Carlo uncertainty.

event time outcome and also for the probability of being censored. Hence, we compare the performance of these methods with respect to both the outcome and the censoring distribution. Again we use the IPA to quantify the predictive performance.

The results are shown in Figures 2 and 3. We see that for most sample sizes, the state learner selected prediction models for both censoring and outcome which have similar or higher IPA compared to the prediction models selected by `survSL`.

8 Prostate cancer study

In this section we use the prostate cancer data of Kattan et al. [2000] to illustrate the use of the state learner in the presence of competing risks. We have introduced the data in Section 7. The data consists of 1,042 patients who are followed from start of followup until tumor recurrence, death without tumor recurrence or end of followup (censored) whatever came first. ... We use the state learner to rank libraries of learners for the cause-specific cumulative hazard functions of tumor recurrence, death without tumor recurrence, and censoring. The libraries of learners each include five learners: the Nelson-Aalen estimator, three Cox regression models (unpenalized, Lasso, Elastic net) each including additive effects of the 5 covariates (Section 7), and a random survival forest. We use the same set of learners to learn the cumulative hazard function of tumor recurrence Λ_1, the cumulative hazard function of death without tumor recurrence Λ_2, and the cumulative hazard function of the conditional censoring distribution Γ.

This gives a library consisting of $5^3 = 125$ learners for the conditional state occupation probability function F defined in equation (5). We use five folds for training and testing the models, and we repeat training and evaluation five times with different splits. The integrated Brier score (defined in Section 4) for all learners are shown in Figure 4. We see that the prediction performance is mostly affected by the choice of learner for the censoring distribution. Several combinations of learners give similar performance as measured by the integrated Brier score, as long as a random forest is used to model the censoring distribution.

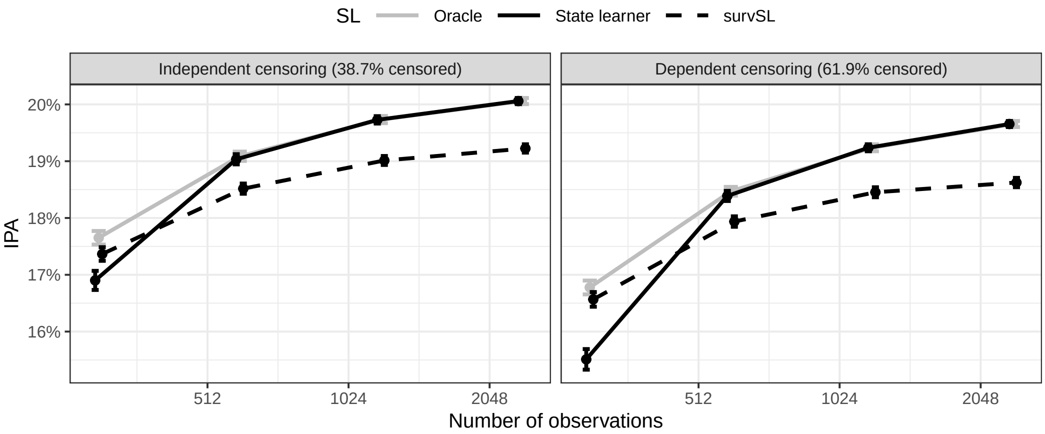

Figure 2: For the risk prediction models of the outcome provided by each of the super learners, the IPA at the fixed time horizon is plotted against sample size. The results are averages across 1000 repetitions and the error bars are used to quantify the Monte Carlo uncertainty.

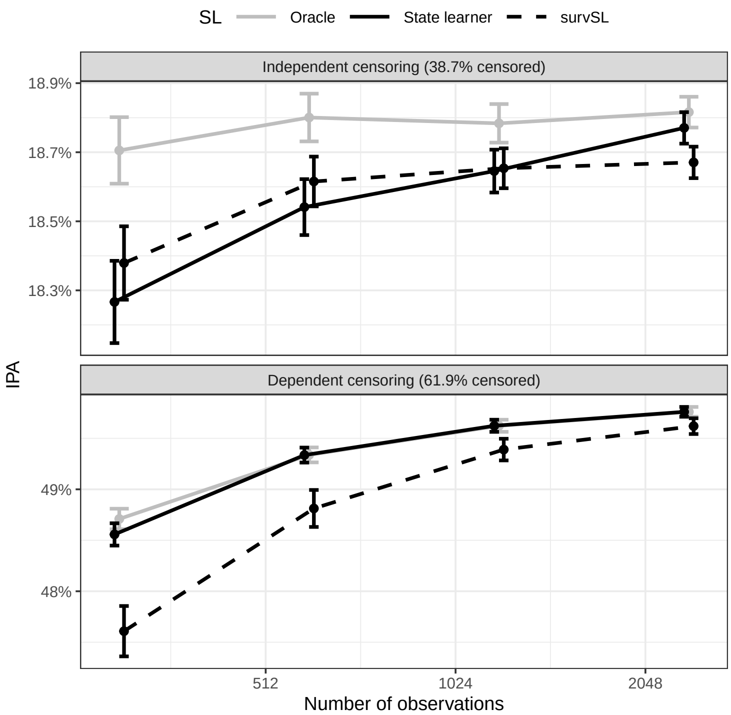

Figure 3: For the risk prediction models of the censoring model provided by each of the super learners, the IPA at the fixed time horizon is plotted against sample size. The results are averages across 1000 repetitions and the error bars are used to quantify the Monte Carlo uncertainty.

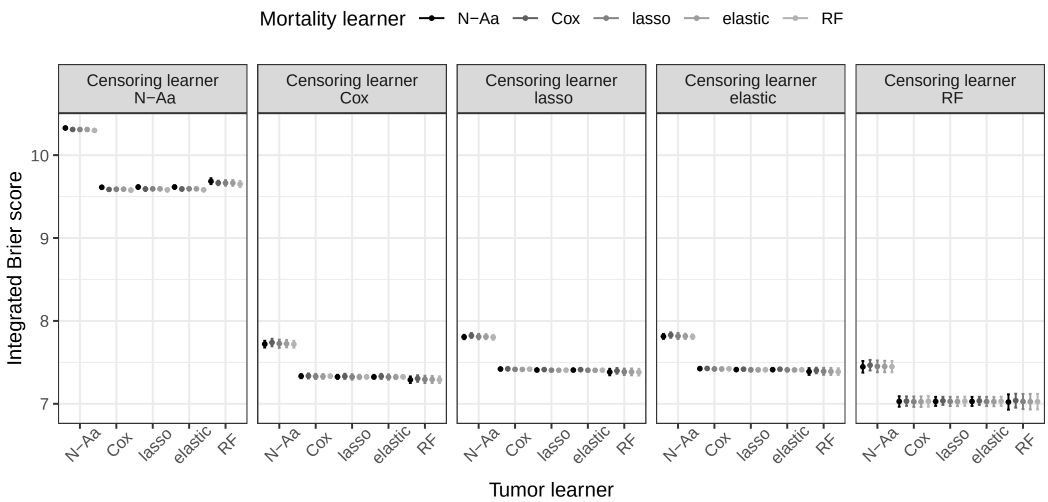

Figure 4: The results of applying the 125 combinations of learners to the prostate cancer data set. The error bars are based on five repetitions using different splits. We refer to learners of Λ_1, Λ_2, and Γ as 'Tumor learner', 'Mortality learner', and 'Censoring learner', respectively.

9 Discussion

The state learner is a new super learner that can be used with right-censored data and competing events. Compared to existing IPCW-based methods, the advantage of the state learner is that it does not depend on a pre-specified estimator of the censoring distribution, but selects one automatically based on a library of learners for the censoring distribution. Furthermore, the state learner neither requires that the cause-specific cumulative hazard functions Λ_j can be written as integrals with respect to Lebesgue measure, nor does it assume a (semi-)parametric formula. In the remainder of this section we discuss the limitations of our proposal and avenues for further research.

A major advantage of the state learner is that the performance of each combination of learners can be estimated without additional nuisance parameters. A potential drawback of our approach is that we are evaluating the loss of the learners on the level of the observed data distribution while the target of the analysis is either the event time distribution, or the censoring distribution, or both. Specifically, the finite sample oracle inequality in Corollary 6.2 concerns the function F, which is a feature of $P \in \mathcal{P}$, while what we are typically interested in is Λ_j or S, which are features of $Q \in \mathcal{Q}$. We emphasise that while the state learner provides us with estimates of Λ_j and Γ based on libraries $\mathcal{A}_j$ and $\mathcal{B}$, the performance of these learners is not assessed directly for their respective target parameters, but only indirectly via the performance of F. For settings without competing risks, our numerical studies suggest that measuring the performance of F also leads to good performance for estimation of S.

Our proposed super learner can be implemented with a broad library of learners and using existing software. Furthermore, while the library $\mathcal{F}(\mathcal{A}_1, \mathcal{A}_2, \mathcal{B})$ consists of $|\mathcal{A}_1||\mathcal{A}_2||\mathcal{B}|$ many learners, we only need to fit $|\mathcal{A}_1| + |\mathcal{A}_2| + |\mathcal{B}|$ many learners in each fold. To evaluate the performance of each learner we need to perform $|\mathcal{A}_1||\mathcal{A}_2||\mathcal{B}|$ many operations to calculate the integrated Brier score in each hold-out sample, one for each combination of the fitted models, but these operations are often negligible compared to fitting the models. Hence the state learner is essentially not more computationally demanding than any procedure that uses super learning to learn Λ_1, Λ_2, and Γ separately. While our proposal is based on constructing the library $\mathcal{F}$ from libraries for learning Λ_1, Λ_2, and Γ, it could also be of interest to consider learners that estimate F directly.

In our numerical studies, we only considered learners of Λ_j and Γ that provide cumulative hazard functions which are piece-wise constant in the time argument. This simplifies the calculation of F as the integrals in equation (6) reduce to sums. When Λ_j or Γ are absolutely continuous in the time argument, calculating F is more involved, but we expect that a good approximation can be achieved by discretisation.

Conflict of interest

The authors declare that they have no conflict of interest.

References

P. K. Andersen, J. P. Klein, and S. Rosthøj. Generalised linear models for correlated pseudo-observations, with applications to multi-state models. *Biometrika*, 2003.

P. K. Andersen, O. Borgan, R. D. Gill, and N. Keiding. *Statistical models based on counting processes*. Springer Science & Business Media, 2012.

J. M. Begun, W. J. Hall, W.-M. Huang, and J. A. Wellner. Information and asymptotic efficiency in parametric-nonparametric models. *The Annals of Statistics*, 11(2):432–452, 1983.

J. Benichou and M. H. Gail. Estimates of absolute cause-specific risk in cohort studies. *Biometrics*, pages 813–826, 1990.

L. Breiman. Stacked regressions. *Machine learning*, 24(1):49–64, 1996.

G. W. Brier et al. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

D. R. Cox. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(2):187–202, 1972.

J. Fan and I. Gijbels. *Local polynomial modelling and its applications*. Routledge, 1996.

M. F. Gensheimer and B. Narasimhan. A scalable discrete-time survival model for neural networks. *PeerJ*, 7:e6257, 2019.

T. A. Gerds. *prodlim: Product-Limit Estimation for Censored Event History Analysis*, 2019. URL `https://CRAN.R-project.org/package=prodlim`. R package version 2019.11.13.

T. A. Gerds and M. W. Kattan. *Medical risk prediction models: with ties to machine learning*. CRC Press, 2021.

T. A. Gerds and M. Schumacher. Consistent estimation of the expected Brier score in general survival models with right-censored event times. *Biometrical Journal*, 48(6):1029–1040, 2006.

T. A. Gerds, M. W. Kattan, M. Schumacher, and C. Yu. Estimating a time-dependent concordance index for survival prediction models with covariate dependent censoring. *Statistics in medicine*, 32(13):2173–2184, 2013.

T. A. Gerds, J. S. Ohlendorff, and B. Ozenne. *riskRegression: Risk Regression Models and Prediction Scores for Survival Analysis with Competing Risks*, 2023. URL https://CRAN.R-project.org/package=riskRegression. R package version 2023.03.22.

R. D. Gill, M. J. van der Laan, and J. M. Robins. Coarsening at random: Characterizations, conjectures, counter-examples. In *Proceedings of the First Seattle Symposium in Biostatistics*, pages 255–294. Springer, 1997.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378, 2007.

M. K. Golmakani and E. C. Polley. Super learner for survival data prediction. *The International Journal of Biostatistics*, 16(2):20190065, 2020.

P. Gonzalez Ginestet, A. Kotalik, D. M. Vock, J. Wolfson, and E. E. Gabriel. Stacked inverse probability of censoring weighted bagging: A case study in the infcarehiv register. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 70(1):51–65, 2021.

E. Graf, C. Schmoor, W. Sauerbrei, and M. Schumacher. Assessment and comparison of prognostic classification schemes for survival data. *Statistics in medicine*, 1999.

X. Han, M. Goldstein, A. Puli, T. Wies, A. Perotte, and R. Ranganath. Inverse-weighted survival games. *Advances in Neural Information Processing Systems*, 34, 2021.

T. Hothorn, P. Bühlmann, S. Dudoit, A. Molinaro, and M. J. van der Laan. Survival ensembles. *Biostatistics*, 7(3):355–373, 2006.

H. Ishwaran and U. Kogalur. *Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)*, 2023. URL https://cran.r-project.org/package=randomForestSRC. R package version 3.2.2.

H. Ishwaran, U. B. Kogalur, E. H. Blackstone, and M. S. Lauer. Random survival forests. *The annals of applied statistics*, 2(3):841–860, 2008.

E. L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53(282):457–481, 1958.

M. W. Kattan and T. A. Gerds. The index of prediction accuracy: an intuitive measure useful for evaluating risk prediction models. *Diagnostic and prognostic research*, 2018.

M. W. Kattan, M. J. Zelefsky, P. A. Kupelian, P. T. Scardino, Z. Fuks, and S. A. Leibel. Pretreatment nomogram for predicting the outcome of three-dimensional conformal radiotherapy in prostate cancer. *Journal of clinical oncology*, 18(19):3352–3359, 2000.

J. L. Katzman, U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger. Deepsurv: personalized treatment recommender system using a Cox proportional hazards deep neural network. *BMC medical research methodology*, 18(1):1–12, 2018.

S. Keles, M. van der Laan, and S. Dudoit. Asymptotically optimal model selection method with right censored outcomes. *Bernoulli*, 10(6):1011–1037, 2004.

H. Kvamme and Ø. Borgan. Continuous and discrete-time survival prediction with neural networks. *Lifetime Data Analysis*, 27(4):710–736, 2021.

C. Lee, W. Zame, J. Yoon, and M. van der Schaar. Deephit: A deep learning approach to survival analysis with competing risks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

D. K. Lee, N. Chen, and H. Ishwaran. Boosted nonparametric hazards with time-dependent covariates. *Annals of Statistics*, 49(4):2101, 2021.

Y. Li, K. S. Xu, and C. K. Reddy. Regularized parametric regression for high-dimensional survival analysis. In *Proceedings of the 2016 SIAM International Conference on Data Mining*, pages 765–773. SIAM, 2016.

U. B. Mogensen and T. A. Gerds. A random forest approach for competing risks based on pseudo-values. *Statistics in medicine*, 32(18):3102–3114, 2013.

A. M. Molinaro, S. Dudoit, and M. J. van der Laan. Tree-based multivariate regression and density estimation with right-censored data. *Journal of Multivariate Analysis*, 90(1): 154–177, 2004.

A. Munch. *Targeted learning with right-censored data*. Phd thesis, University of Copenhagen, 2024. URL https://publichealth.ku.dk/about-the-department/biostat/phd-theses/2023_munch.pdf.

B. Ozenne, A. L. Sørensen, T. Scheike, C. Torp-Pedersen, and T. A. Gerds. riskregression: Predicting the risk of an event using Cox regression models. *R Journal*, 9(2):440–460, 2017.

E. C. Polley and M. J. van der Laan. Super learning for right-censored data. In M. J. van der Laan and S. Rose, editors, *Targeted Learning: Causal Inference for Observational and Experimental Data*, pages 249–258. Springer, 2011.

H. C. Rytgaard and M. J. van der Laan. Targeted maximum likelihood estimation for causal inference in survival and competing risks analysis. *Lifetime Data Analysis*, pages 1–30, 2022.

H. C. Rytgaard, F. Eriksson, and M. J. van der Laan. Estimation of time-specific intervention effects on continuously distributed time-to-event outcomes by targeted maximum likelihood estimation. *Biometrics*, 2021.

M. C. Sachs, A. Discacciati, Å. H. Everhov, O. Olén, and E. E. Gabriel. Ensemble prediction of time-to-event outcomes with competing risks: A case-study of surgical complications in Crohn's disease. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 68(5):1431–1446, 2019.

J. A. Steingrimsson, L. Diao, and R. L. Strawderman. Censoring unbiased regression trees and ensembles. *Journal of the American Statistical Association*, 2019.

T. M. Therneau. *A Package for Survival Analysis in R*, 2022. URL https://CRAN.R-project.org/package=survival. R package version 3.4-0.

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, 2003.

M. J. van der Laan and S. Rose. *Targeted learning: causal inference for observational and experimental data*. Springer Science & Business Media, 2011.

M. J. van der Laan, E. C. Polley, and A. E. Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6(1), 2007.

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

P. J. Verweij and H. C. van Houwelingen. Cross-validation in survival analysis. *Statistics in medicine*, 12(24):2305–2314, 1993.

T. Westling, A. Luedtke, P. Gilbert, and M. Carone. Inference for treatment-specific survival curves using machine learning. *arXiv preprint arXiv:2106.06602*, 2021.

D. H. Wolpert. Stacked generalization. *Neural networks*, 5(2):241–259, 1992.

J. Yao, X. Zhu, F. Zhu, and J. Huang. Deep correlational learning for survival prediction from multi-modality data. In *International conference on medical image computing and computer-assisted intervention*, pages 406–414. Springer, 2017.

A Proofs

This section contains proofs of the results stated in the paper. Section A.1 contains a proof of the consistency result from Section 6.1, and Section A.2 contains proofs of the oracle inequalities from Section 6.2.

A.1 Consistency

Define $\bar{B}_{\tau,0}(F,o) = \bar{B}_\tau(F,o) - \bar{B}_\tau(F_0,o)$ and $R_0(F) = P_0[\bar{B}_{\tau,0}(F,\cdot)]$, where the integrated Brier score $\bar{B}_\tau$ was defined in Section 4.

Lemma A.1. *$R_0(F) = \|F - F_0\|_{P_0}^2$, where $\|\cdot\|_{P_0}$ is defined in equation (9).*

Proof. For any $t \in [0,\tau]$ and $j \in \{-1,0,1,2\}$ we have

$$
\begin{aligned}
&\mathbb{E}_{P_0}\left[(F(t,j,X) - \mathbb{1}\{\eta(t) = j\})^2\right] \\
&= \mathbb{E}_{P_0}\left[(F(t,j,X) - F_0(t,j,X) + F_0(t,j,X) - \mathbb{1}\{\eta(t) = j\})^2\right] \\
&= \mathbb{E}_{P_0}\left[(F(t,j,X) - F_0(t,j,X))^2\right] + \mathbb{E}_{P_0}\left[(F_0(t,j,X) - \mathbb{1}\{\eta(t) = j\})^2\right] \\
&\quad + 2\,\mathbb{E}_{P_0}\left[(F(t,j,X) - F_0(t,j,X))(F_0(t,j,X) - \mathbb{1}\{\eta(t) = j\})\right] \\
&= \mathbb{E}_{P_0}\left[(F(t,j,X) - F_0(t,j,X))^2\right] + \mathbb{E}_{P_0}\left[(F_0(t,j,X) - \mathbb{1}\{\eta(t) = j\})^2\right],
\end{aligned}
$$

where the last equality follows from the tower property. Hence, using Fubini, we have

$$
P[\bar{B}_\tau(F,\cdot)] = \|F - F_0\|_{P_0}^2 + P_0[\bar{B}_\tau(F_0,\cdot)].
$$

□

Proof of Proposition 6.1. The result follows from Lemma A.1. □

A.2 Oracle inequalities

Recall that we use $\mathcal{F}_n$ to denote a library of learners for the function F, and that $\hat{\varphi}$ and $\tilde{\varphi}$ denotes, respectively, the discrete super learner and the oracle learner for the library $\mathcal{F}_n$, c.f., Section 4.

Proof of Corollary 6.2. First note that minimising the loss $\bar{B}_\tau$ is equivalent to minimising the loss $\bar{B}_{\tau,0}$, so the discrete super learner and oracle according to $\bar{B}_\tau$ and $\bar{B}_{\tau,0}$ are identical.

By Lemma A.1, $R_0(F) \geq 0$ for any $F \in \mathcal{H}_{\mathcal{P}}$, and so using Theorem 2.3 from [van der Vaart et al., 2006] with $p = 1$, we have that for all $\delta > 0$,

$$\begin{aligned}
&\mathbb{E}_{P_0}\left[R_0(\hat{\varphi}_n(\mathcal{D}_n^{-k}))\right] \\
&\quad \leq (1+2\delta)\,\mathbb{E}_{P_0}\left[R_0(\tilde{\varphi}_n(\mathcal{D}_n^{-k}))\right] \\
&\qquad + (1+\delta)\frac{16K}{n}\log(1+|\mathcal{F}_n|)\sup_{F\in\mathcal{H}_{\mathcal{P}}}\left\{M(F) + \frac{v(F)}{R_0(F)}\left(\frac{1}{\delta}+1\right)\right\}
\end{aligned}$$

where for each $F \in \mathcal{H}_{\mathcal{P}}$, $(M(F), v(F))$ is some Bernstein pair for the function $o \mapsto \bar{B}_{\tau,0}(F, o)$. As $\bar{B}_{\tau,0}(F, \cdot)$ is uniformly bounded by τ for any $F \in \mathcal{H}_{\mathcal{P}}$, it follows from section 8.1 in [van der Vaart et al., 2006] that $(\tau, 1.5P_0[\bar{B}_{\tau,0}(F,\cdot)^2])$ is a Bernstein pair for $\bar{B}_{\tau,0}(F,\cdot)$. Now, for any $a, b, c \in \mathbb{R}$ we have

$$\begin{aligned}
(a-c)^2 - (b-c)^2 &= (a-b+b-c)^2 - (b-c)^2 \\
&= (a-b)^2 + (b-c)^2 + 2(b-c)(a-b) - (b-c)^2 \\
&= (a-b)\left\{(a-b) + 2(b-c)\right\} \\
&= (a-b)\left\{a+b-2c\right\},
\end{aligned}$$

so using this with $a = F(t,j,x)$, $b = F_0(t,j,x)$, and $c = \mathbb{1}\{\eta(t) = j\}$, we have by Jensen's inequality

$$\begin{aligned}
&P_0[\bar{B}_{\tau,0}(F,\cdot)^2] \\
&\leq 2\tau\,\mathbb{E}_{P_0}\left[\sum_{j=-1}^{2}\int_0^\tau \left\{(F(t,j,X) - \mathbb{1}\{\eta(t)=j\})^2 - (F_0(t,j,X) - \mathbb{1}\{\eta(t)=j\})^2\right\}^2 \mathrm{d}t\right] \\
&= 2\tau\,\mathbb{E}_{P_0}\left[\sum_{j=-1}^{2}\int_0^\tau (F(t,j,X) - F_0(t,j,X))^2 \right. \\
&\qquad\qquad \left. \times \left\{F(t,j,X) + F_0(t,j,X) - 2\mathbb{1}\{\eta(t)=j\}\right\}^2 \mathrm{d}t\right] \\
&\leq 8\tau\,\mathbb{E}_{P_0}\left[\sum_{j=-1}^{2}\int_0^\tau (F(t,j,X) - F_0(t,j,X))^2\,\mathrm{d}t\right]. \\
&= 8\tau\|F - F_0\|_{P_0}^2.
\end{aligned}$$

Thus when $v(F) = 1.5P_0[\bar{B}_{\tau,0}(F,\cdot)^2]$ we have by Lemma A.1

$$\frac{v(F)}{R_0(F)} = 1.5\frac{P_0[\bar{B}_{\tau,0}(F,\cdot)^2]}{P_0[\bar{B}_{\tau,0}(F,\cdot)]} \leq 12\tau,$$

and so using the Bernstein pairs $(\tau, 1.5P_0[\bar{B}_{\tau,0}(F,\cdot)^2])$ we have

$$\sup_{F\in\mathcal{H}_{\mathcal{P}}}\left\{M(F) + \frac{v(F)}{R_0(F)}\left(\frac{1}{\delta}+1\right)\right\} \leq \tau\left(13 + \frac{12}{\delta}\right),$$

For all $\delta > 0$ we thus have

$$\begin{aligned}
\mathbb{E}_{P_0}\left[R_0(\hat{\varphi}_n(\mathcal{D}_n^{-k}))\right] \leq &(1+2\delta)\,\mathbb{E}_{P_0}\left[R_0(\tilde{\varphi}_n(\mathcal{D}_n^{-k}))\right] \\
&+ (1+\delta)\log(1+|\mathcal{F}_n|)\tau\frac{16K}{n}\left(13 + \frac{12}{\delta}\right),
\end{aligned}$$

and then the final result follows from Lemma A.1. □

Proof of Corollary 6.3. By definition of the oracle and Lemma A.1,

$$\mathbb{E}_{P_0}\left[\|\tilde{\varphi}_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right] \leq \mathbb{E}_{P_0}\left[\|\varphi_n(\mathcal{D}_n^{-k}) - F_0\|_{P_0}^2\right]$$

for all $n \in \mathbb{N}$. The results then follows from Corollary 6.2. $\square$