Supplementary material: The state learner – a super learner for right-censored data

Anders Munch[1,*] and Thomas A. Gerds[1]

[1]Section of Biostatistics, University of Copenhagen

[*]Address for correspondence: a.munch@sund.ku.dk

December 29, 2024

This document contains proofs of the results stated in the main paper. Section S1 contains a proof of the consistency result from Section 5.1 of the main paper; Section S2 contains proofs of the oracle inequalities from Section 5.2 of the main paper; and Section S3 demonstrates transience of the second order remainder structure stated in Section 5.3 of the main paper.

We start by recalling some notation. We use F to denote a function that describes the conditional state occupation probabilities of the observed data, as defined in equation (5) of the main paper. We denote the integrated Brier score as $\bar{B}_\tau(F, O) = \int_0^\tau B_t(F, O)\,\mathrm{d}t$, where B_t is defined

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

Finally, recall that we let $\mathcal{H}_\mathcal{P}$ denote the function space consisting of all conditional state occupation probability functions for some measure $P \in \mathcal{P}$, and that we equipped this space with the norm

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

for some fixed measure $P_0 \in \mathcal{P}$. We use F_0 to denote the conditional state occupation probability function associated with the measure P_0.

S1 Consistency

Define $\bar{B}_{\tau,0}(F,o) = \bar{B}_{\tau}(F,o) - \bar{B}_{\tau}(F_0,o)$ and $R_0(F) = P_0[\bar{B}_{\tau,0}(F,\cdot)]$.

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

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 1 of the main paper. The result follows from Lemma 1. □

S2 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 of the main paper.

Proof of Corollary 1 of the main paper. 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 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 $\delta > 0$,

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

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

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

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

Proof of Corollary 2 of the main paper. By definition of the oracle and Lemma 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 1 of the main paper. □

S3 Transience of the second order remainder structure

Recall that we let Λ_1 denote the conditional cumulative hazard function for one of the event times of interest and Γ the conditional cumulative hazard function for censoring, c.f., Section 2 of the main paper.

Proof of Proposition 2 of the main paper. For notational convenience we suppress X in the following. The final result can be obtained by adding the argument X to all functions and averaging. We use the relations from equation (8) in the main paper to write

$$
\begin{aligned}
&\int_0^\tau w(s)\left\{\Gamma(s)-\hat{\Gamma}_n(s)\right\}[\Lambda_1-\hat{\Lambda}_{1,n}](\mathrm{d}s)\\
&=\int_0^\tau w(s)\left\{\int_0^s \frac{F(\mathrm{d}u,-1)}{F(u-,0)}-\int_0^s \frac{\hat{F}_n(\mathrm{d}u,-1)}{\hat{F}_n(u-,0)}-\right\}\left[\frac{F(\mathrm{d}s,1)}{F(s-,0)}-\frac{\hat{F}_n(\mathrm{d}s,1)}{\hat{F}_n(s-,0)}\right]\\
&=\int_0^\tau w(s)\Bigg\{\int_0^s \left(\frac{1}{F(u-,0)}-\frac{1}{\hat{F}_n(u-,0)}\right)F(\mathrm{d}u,-1)\\
&\qquad\qquad +\int_0^s \frac{1}{\hat{F}_n(u-,0)}\left[F(\mathrm{d}u,-1)-\hat{F}_n(\mathrm{d}u,-1)\right]\Bigg\}\\
&\qquad\quad \times\left[\left(\frac{1}{F(s-,0)}-\frac{1}{\hat{F}_n(s-,0)}\right)F(\mathrm{d}s,1)+\frac{1}{\hat{F}_n(s-,0)}\left(F(\mathrm{d}s,1)-\hat{F}_n(\mathrm{d}s,1)\right)\right]\\
&=\int_0^\tau\int_0^s w(s)\left(\frac{1}{F(u-,0)}-\frac{1}{\hat{F}_n(u-,0)}\right)\left(\frac{1}{F(s-,0)}-\frac{1}{\hat{F}_n(s-,0)}\right)F(\mathrm{d}u,-1)F(\mathrm{d}s,1)\\
&\quad+\int_0^\tau\int_0^s w(s)\left(\frac{1}{F(u-,0)}-\frac{1}{\hat{F}_n(u-,0)}\right)\frac{F(\mathrm{d}u,-1)}{\hat{F}_n(u-,0)}\left(F(\mathrm{d}s,1)-\hat{F}_n(\mathrm{d}s,1)\right)\\
&\quad+\int_0^\tau\int_0^s \frac{w(s)}{\hat{F}_n(u-,0)}\left[F(\mathrm{d}u,-1)-\hat{F}_n(\mathrm{d}u,-1)\right]\left(\frac{1}{F(s-,0)}-\frac{1}{\hat{F}_n(s-,0)}\right)F(\mathrm{d}s,1)\\
&\quad+\int_0^\tau\int_0^s \frac{w(s)}{\hat{F}_n(u-,0)}\left[F(\mathrm{d}u,-1)-\hat{F}_n(\mathrm{d}u,-1)\right]\frac{1}{\hat{F}_n(s-,0)}\left(F(\mathrm{d}s,1)-\hat{F}_n(\mathrm{d}s,1)\right).
\end{aligned}
$$

Consider the first term on the right hand side. By the mean value theorem,

$$
\frac{1}{F(t-,0)}-\frac{1}{\hat{F}_n(t-,0)}=\frac{-1}{\tilde{r}_n(t)^2}\left[F(t-,0)-\hat{F}_n(t-,0)\right],
$$

where $\tilde{r}_n(t)$ is some value between $\hat{F}(t-,0)$ and $\hat{F}_n(t-,0)$. Letting $w_n^*(t)=-\tilde{r}_n(t)^{-2}$, we may write

$$\begin{aligned}
&\int_0^\tau \int_0^s w(s) \left(\frac{1}{F(u-,0)} - \frac{1}{\hat{F}_n(u-,0)} \right) \left(\frac{1}{F(s-,0)} - \frac{1}{\hat{F}_n(s-,0)} \right) F(\mathrm{d}u,-1)F(\mathrm{d}s,1) \\
&= \int_0^\tau \int_0^s w(s) w_n^*(u) \left(F(u-,0) - \hat{F}_n(u-,0) \right) \\
&\qquad\qquad \times w_n^*(s) \left(F(s-,0) - \hat{F}_n(s-,0) \right) F(\mathrm{d}u,-1)F(\mathrm{d}s,1) \\
&= \int_0^\tau \int_0^s w_n^a(s,u) \left(F(u-,0) - \hat{F}_n(u-,0) \right) \left(F(s-,0) - \hat{F}_n(s-,0) \right) F(\mathrm{d}u,-1)F(\mathrm{d}s,1),
\end{aligned}$$

where we have defined $w_n^a(s,u) = w(s)w_n^*(s)w_n^*(u)$. By the assumption that $F(\cdot,0)$ and $F_n(\cdot,0)$ are uniformly bounded away from zero on $[0,\tau]$, it follows that w_n^* is uniformly bounded, and hence $w_n^a(s,u)$ is also uniformly bounded, because w was assumed uniformly bounded. The same approach can be applied to the three remaining terms which gives the result. □

References

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.