The state learner

Anders Munch
joint work with Thomas Gerds

October 9, 2023

Outline

Super learning (aka cross-validation, stacked regression, ...)

Consider estimating a conditional mean $f(x) = \mathbb{E}[Y \mid X = x]$ based on data $\mathcal{D}_n = \{O_1, \ldots, O_n\}$, where $O_i = (X_i, Y_i)$ are iid. observations.

Learner algorithm a that produces estimates, $\mathcal{D}_n \mapsto a(\mathcal{D}_n) = \hat{f}$

Library collection of learners, $\mathcal{A} = \{a_1, a_2, \ldots, a_M\}$

Loss function $L(a, O)$, with $O = (X, Y)$

Stand-alone prediction E.g., estimating conditional survival probabilities. Can be used to select hyperparameters or choose between a collection of (parametric) models.

Nuisance parameter estimation E.g., using TMLE to estimate the average treatment effect where the outcome and propensity models are estimated using super learners.

Right-censored data

X vector of baseline covariates

$\tilde{T}$ censored time to event variable, $\tilde{T} = \min(T, C),\ T, C \geq 0$

Δ binary event indicator, $\Delta = \mathbb{1}\{T \leq C\}$

Parameter of interest is a feature of Q where we use

$(X, T) \sim Q$ The distribution of interest

$(X, \tilde{T}, \Delta) \sim P$ The observed data distribution

We use Λ and Γ , respectively, to denote the conditional cumulative hazard function for T and C, i.e.,

$$\Lambda(\mathrm{d}t \mid x) = Q(T \in \mathrm{d}t \mid T \geq t, X = x).$$

We assume $T \perp\!\!\!\perp C \mid X$. Implies that Λ and Γ are identifiable from P.

Super learning with right-censored data

Use super learning to estimate Λ or $S(t \mid X = x) = Q(T > t \mid X = x)$.
Challenge is to handle censoring.

Existing approaches

Negative log-likelihood loss function

Requires modeling a Lebesgue hazard function which is incompatible with many common estimator in survival analysis (e.g., Kaplan-Meier, semi-parametric Cox models, random survival forests).

Pseudo-observations

Require pre-specification of the censoring estimator.

Inverse probability of censoring weighting

Require pre-specification of the censoring estimator.
To avoid this, it has recently been suggested to iterate between estimation of Λ and Γ. No theoretical guarantees for this procedure.

Idea: An observed multi-state system

Modeling the conditional state-occupation probabilities of the *observed* data

Learners for the conditional state-occupation probabilities

Build library of learner of F using learner of Λ and Γ.

Some theoretical results

Unique minimizer

Oracle inequality

Some simulations

Competing events

Some limitations

Target parameters

References