Efficient Estimation of the Maximal Association between Multiple Predictors and a Survival Outcome

Alex Luedtke (♙)

Joint work with:

Tzu-Jung Huang (♗)
Ian McKeague (♘)

♙ : Department of Statistics, University of Washington
♗ : Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Center
♘ : Department of Biostatistics, Columbia University

Acknowledgements

This research was supported by the:

- **NIH** through award numbers DP2-LM013340 and 1R01 AG062401;
- **NSF** through award number DMS-2112938.

The work that I speak about today is based on the following preprint:

- Huang, Tzu-Jung, Alex Luedtke, and Ian W. McKeague. "Efficient Estimation of the Maximal Association between Multiple Predictors and a Survival Outcome." arXiv preprint arXiv:2112.10996 (2021).

Data structure: right-censored survival time with high-dimensional feature

- T: time-to-event outcome
- $\boldsymbol{U} = (U_1, \ldots, U_p)$: p-dimensional vector of features

T may by censored by an independent variable C, in which case we observe

$$X = \min\{T, C\} \quad \text{and} \quad \Delta = I\{T \leq C\}$$

Objective

Our objectives are to develop a **confidence interval** for

$$\max_k \Psi_k(P) = \max_k \frac{\mathrm{Cov}_P(U_k, T)}{\mathrm{Var}_P(U_k)}$$

and a **hypothesis test** for

$$H_0 : \max_k \Psi_k(P) = 0 \quad \text{vs.} \quad H_1 : \text{ not } H_0.$$

Examples

Virology:

- potency of a monoclonal antibody is assessed in terms of a survival outcome
- it is important to identify associations with patterns of viral gene expression

Cancer genomics:

- identifying relationships between patients' gene expression and their survival times

An easier problem: inferring the association between T and one feature

To develop some tools that we'll need later, we first study a simpler setting.

For now, for a **fixed feature** k, our goal is infer about

$$\Psi_k(P) = \frac{\mathrm{Cov}_P(U_k, T)}{\mathrm{Var}_P(U_k)}.$$

In what follows we develop an **efficient one-step estimator** for this quantity.

Nuisance function estimation in the easier problem

To construct an efficient one-step estimator, we require estimators of:

(i) **marginal distribution of U**
- We estimate with the empirical distribution.

(ii) **censoring distribution**, namely $G(t) = P(C \geq t)$
- We estimate with the Kaplan-Meier estimator.

(iii) **conditional mean residual life function**
$(u, s, k) \mapsto E[T - s | U_k = u, T \geq s]$
- We allow for the use of an **arbitrary consistent estimator** of this quantity.

We let $\widehat{P}_n$ be any distribution whose nuisances (i), (ii), and (iii) correspond to the estimates described above.

One-step estimation in the easier problem

The efficient one-step estimator for $\Psi_k(P)$ takes the form

$$\hat{\psi}_k = \Psi_k(\widehat{P}_n) + \frac{1}{n}\sum_{i=1}^{n} \mathrm{IF}_k(U_{k,i}, X_i, \Delta_i \mid \widehat{P}_n),$$

where $\mathrm{IF}_k(\cdot \mid \widehat{P}_n)$ denotes the **efficient influence function** of Ψ at $\widehat{P}_n$ relative to a nonparametric model.

- We derive the form of IF_k in our paper.

Under regularity conditions,

$$n^{1/2}\left[\hat{\psi}_k - \Psi_k(P)\right] \rightsquigarrow N(0, \sigma_k^2),$$

which facilitates the construction of **Wald-type confidence intervals**.

Back to the harder problem: inferring the maximal association between T and one of p features

We now return to the problem of constructing a confidence interval for

$$\max_k \Psi_k(P) = \max_k \frac{\mathrm{Cov}_P(U_k, T)}{\mathrm{Var}_P(U_k)}.$$

To do this, we must incorporate the **selection of the maximally associated predictor** into our inferential scheme.

- Indeed, the estimator $\max_k \hat{\psi}_k$ will generally be **positively biased asymptotically**, which makes using it as a basis for inference difficult.

A simple variant of our approach: sample splitting

Before presenting our approach in full generality, I'll present a simple version of it based on **sample splitting**.

1. For each k, let
   - $\hat{\psi}_k^{(1)}$ and $\hat{\psi}_k^{(2)}$ denote a **one-step estimator** for $\Psi_k(P)$ based on the first and second half of the data, respectively.
   - $\hat{\sigma}_k$ be an estimator of the limiting standard deviation σ_k of $\hat{\psi}_k^{(2)}$.
2. **Estimate a maximizing index** of $\Psi_k(P)$ via

$$\hat{k} \in \operatorname*{argmax}_k \hat{\psi}_k^{(1)}.$$

3. Return as **confidence interval**

$$\hat{\psi}_{\hat{k}}^{(2)} \pm 1.96 \frac{\hat{\sigma}_{\hat{k}}}{\sqrt{n/2}}.$$

A simple variant of our approach: sample splitting

Establishing the asymptotic coverage of our confidence interval relies on showing:

- $\hat{k}$ is **nearly a maximizing index** of $\Psi_k(P)$, in the sense that

$$\Psi_{\hat{k}}(P) = \max_k \Psi_k(P) + o_p(n^{1/2});$$

- the **uniformity of the linearization argument** used to justify the validity of a one-step estimator for $\Psi_k(P)$ over $k \in \{1, 2, \ldots, p\}$.

We can justify both of these properties under the condition that $\log(p)/n^{1/4} \to 0$ with sample size, which allows p to grow rapidly with n.

- The most challenging part of providing this guarantee involves finding an exponential tail bound for a "remainder term" involving **martingale integrals with unpredictable integrands** that change with n.

A simple variant of our approach: sample splitting

Drawback of sample-splitting:

- Width of confidence interval is determined by only **half of the observations**.

Cannot overcome this drawback via cross-fitting.

- The **non-regularity** of the parameter $\max_k \Psi_k(\cdot)$ would render the resulting confidence interval invalid.
- This non-regularity owes to the **non-smooth maximization operation**.

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- For $v = 2$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- For $v = 2$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- For $v = 2$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- For $v = 3$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

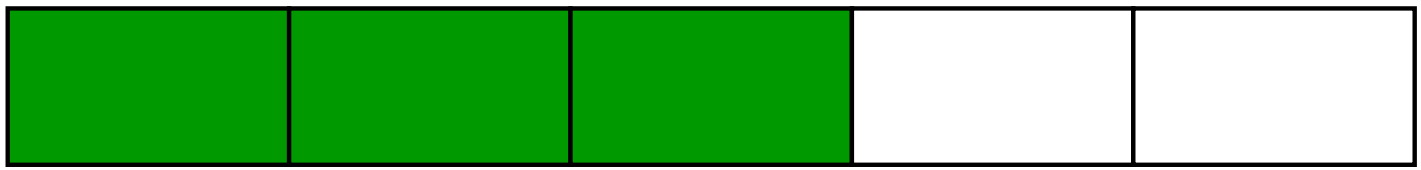
Evaluate one-step estimator

- For $v = 3$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:



Estimate a maximizing index and nuisance functions

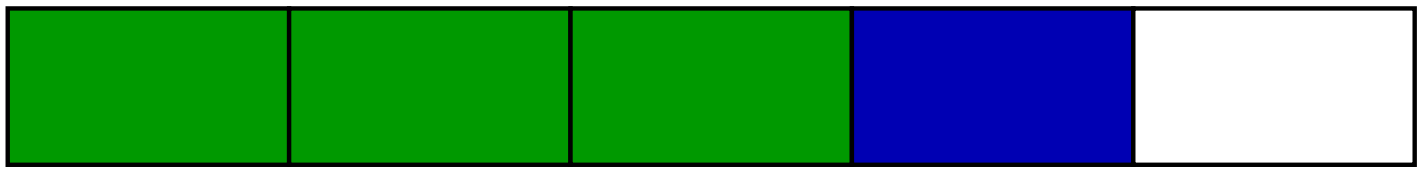
Evaluate one-step estimator

- For $v = 4$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:



Estimate a maximizing index and nuisance functions

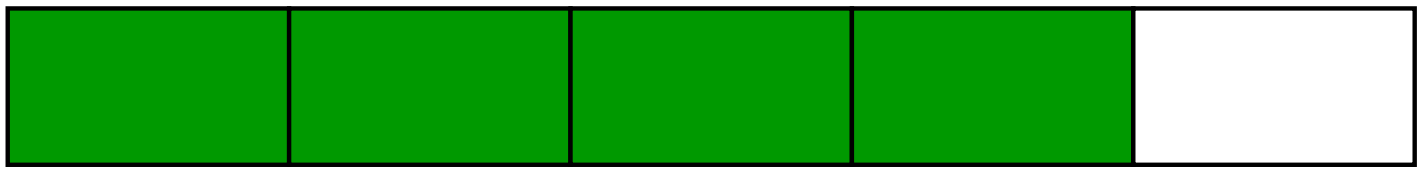
Evaluate one-step estimator

- For $v = 4$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:



Estimate a maximizing index and nuisance functions

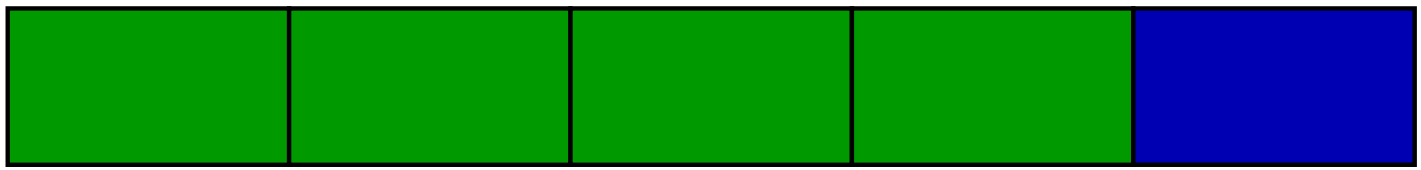
Evaluate one-step estimator

- For $v = 5$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:



Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- For $v = 5$, compute estimator:

$$\hat{\psi}^{(v)} = \Psi_{\hat{k}^{(v-1)}}(\hat{P}^{(v-1)}) + \frac{1}{|\text{Chunk } v|} \sum_{i \in \text{Chunk } v} \text{IF}_{\hat{k}^{(v-1)}}(\boldsymbol{U}_i, X_i, \Delta_i \mid \hat{P}^{(v-1)})$$

Overview of estimator

- Split data into V chunks. Here we let $V = 5$:

Estimate a maximizing index and nuisance functions

Evaluate one-step estimator

- Now have four estimates $\hat{\psi}^{(2)}, \ldots, \hat{\psi}^{(5)}$

Overview of estimator

- Final estimate an average of the four chunk-based estimates:

$$\hat{\psi}_{\text{final}} = \sum_{v=2}^{V} w^{(v-1)} \hat{\psi}^{(v)},$$

where $w^{(v-1)}$ are convex weights inversely proportional to the estimated asymptotic standards deviation of $\hat{\psi}^{(v)}$

- For appropriately defined $\bar{\sigma}_n$ and under similar conditions to those used in the sample-splitting special case, a **martingale CLT** yields that

$$\frac{\sqrt{n'}\left[\hat{\psi}_{\text{final}} - \max_k \Psi_k(P)\right]}{\bar{\sigma}_n} \xrightarrow{d} \text{Normal}(0, 1),$$

where $n' = \frac{V-1}{V} \times n$, which facilitates the construction of confidence intervals.

Weakness of our estimator and a simple, but surprisingly effective, fix

As presented thus far, our estimator depends on the **order of the data** used to define the chunks.

Though it is tempting to remove this dependence by **averaging over many orderings**, it is unclear how to make inference based on this estimator.

- The problem is that this averaging scheme **breaks the martingale structure** that we use to justify our approach.

We instead consider an alternative testing strategy:

1. Define 10 different orderings of the data.
2. Compute our estimator on each of these orderings.
3. Report minimal, **Bonferroni-corrected p-value** (correcting for 10 tests).

A similar approach can be used to construct confidence intervals.

Setting

Survival times are generated under one of the following scenarios:

- **Model N** (no signal feature): $T = \varepsilon$;
- **Model A1** (1 signal feature): $T = U_1/4 + \varepsilon$;
- **Model A2** (10 signal features): $T = \sum_{k=1}^{p} \beta_k U_k + \varepsilon$ with $\beta_1 = \ldots = \beta_5 = 0.15$, $\beta_6 = \ldots = \beta_{10} = -0.1$, $\beta_k = 0$ for $k \geq 11$.

The features are such that

$$\boldsymbol{U} \sim N\left(\mathbf{0}, \begin{bmatrix} 1 & 0.25 & 0.25 & \ldots & 0.25 \\ 0.25 & 1 & 0.25 & \ldots & 0.25 \\ 0.25 & 0.25 & 1 & \ldots & 0.25 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.25 & 0.25 & 0.25 & \ldots & 1 \end{bmatrix}\right)$$

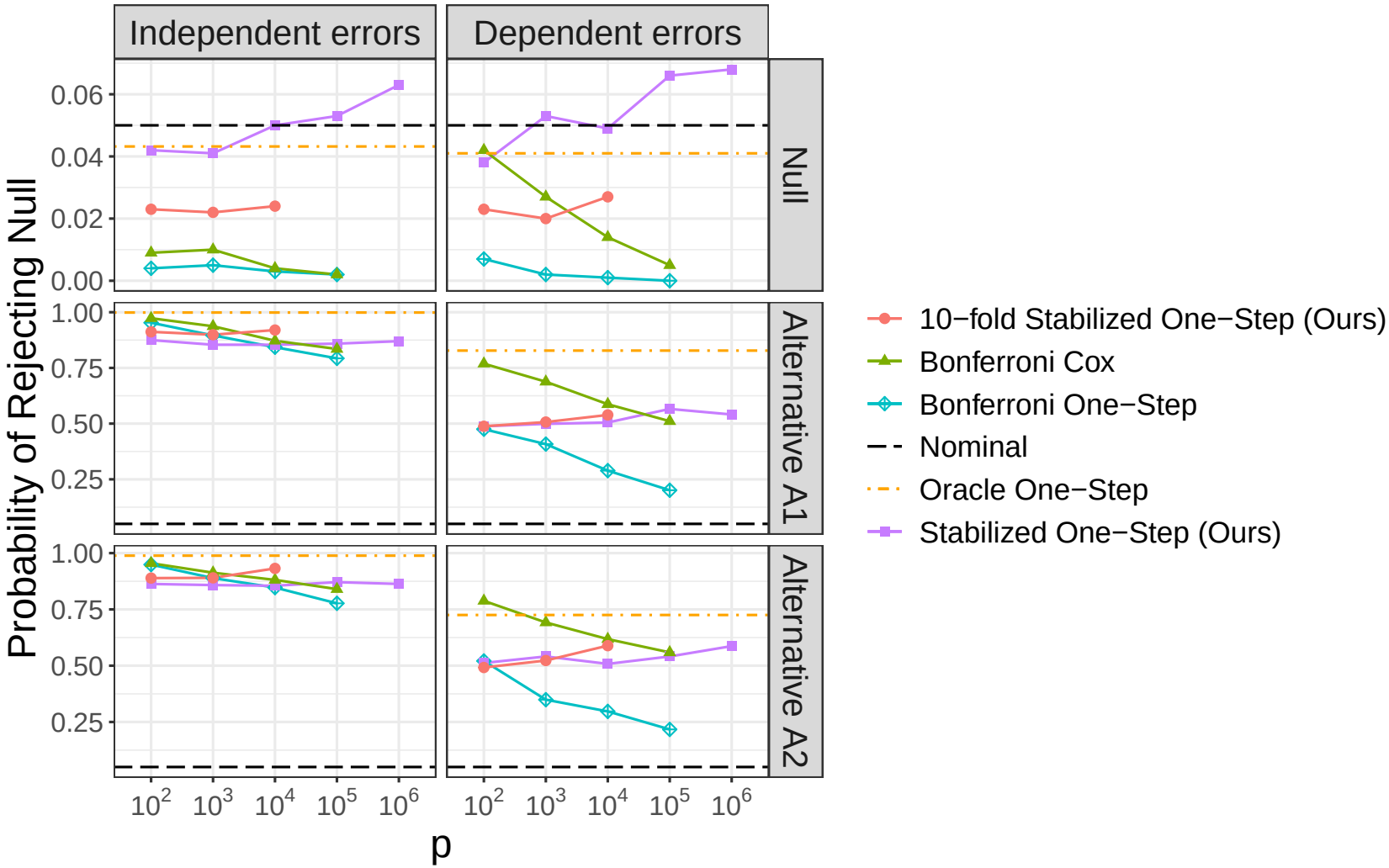
The errors are either

- Independent: $\varepsilon \mid \boldsymbol{U} \sim N(0, 1)$, or
- Dependent: $\varepsilon \mid \boldsymbol{U} \sim N(0, 0.7[|U_1| + 0.7])$.

We study the performance of a test of

$$H_0 : \max_k \Psi_k(P) = 0 \quad \text{vs.} \quad H_1 : \text{ not } H_0.$$

Results ($n = 500$)



Concluding remarks

We've described a means to evaluate the maximal association of a feature with a right-censored outcome.

- **Number of features can be large relative to sample size**, nearly on the order of $\exp(n^{1/4})$.

While we characterized association via a slope parameter, **other parameters could also be considered**.

- The framework generalizes naturally to other association measures **provided an asymptotically linear estimator is available** in the univariate setting.
- However, a new analysis would be needed to describe the extent to which p can grow with n.

Thank you!

Example from HIV prevention studies

Goal: assess whether there are **viral characteristics that predict the potency of a monoclonal antibody** for preventing HIV replication

- Each observation is a pseudovirus
- T = the concentration needed to achieve a 50% reduction of the (in vitro) rate of viral replication (termed the IC_{50})
- For highly-resistant viruses, viral replication may not be reduced by 50% even at the maximal observed concentration (C), so that T is right-censored.
- $\boldsymbol{U}$ consists of characteristics of the virus

We looked at data on the neutralization of 293 HIV pseudo-viruses, generated by the HIV Vaccine Trials Network [2].

Results

Table: Results of applying the Bonferroni one-step estimator and the stabilized one-step estimator to data on Subtype C. The data consist of 293 pseudoviruses, 12950 binary predictors, and 153 count predictors.

Method	Binary effects		Count effects	
	95% CI	p-value	95% CI	p-value
Bonferroni Cox	–	< 0.001	–	0.34
Bonferroni One-Step	–	< 0.001	–	0.91
Stabilized One-Step	$(10.4, 23.1)$	< 0.001	$(3.3, 5.0)$	< 0.001