Estimating treatment effects with optimal inverse probability weighting

RSS 2017 Conference - Glasgow

Michele Santacatterina
michele.santacatterina@ki.se

Matteo Bottai
matteo.bottai@ki.se

Unit of Biostatistics, Karolinska Institutet

September 6th 2017

```
> head(hmohiv)
  ID time age drug censor    entdate     enddate
1  1    5  46    0      1 5/15/1990  10/14/1990
2  2    6  35    1      0 9/19/1989   3/20/1990
3  3    8  30    1      1 4/21/1991  12/20/1991
4  4    3  30    1      1  1/3/1991    4/4/1991
5  5   22  36    0      1 9/18/1989   7/19/1991
6  6    1  32    1      0 3/18/1991   4/17/1991

> m1 <- coxph(Surv(time,censor)~drug,data=hmohiv)
> coef(summary(m1))[3]
[1] 0.2418138

> mw <- coxph(Surv(time,censor)~drug,data=hmohiv, weights = w)
> coef(summary(mw))[3]
[1] 8.123295

> summary(w)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.00    0.00    0.00    1.00    0.00   79.34
```

Objective

1. We have target inverse probability weights, $\hat{w}^*$.
2. The weighted estimate has large variance.
3. We estimate the weights closest to $\hat{w}^*$ within a variance constraint.

We propose a general method to estimate optimal inverse probability weights based on the solution of a nonlinear constrained optimization problem [1].

Introduction

Inverse probability weighting is frequently used in the medical literature, despite its being prone to give highly inefficient inference in the presence of outlying probability weights.

This issue can be overcome by

- weight trimming, or truncation, which consists of replacing outlying weights with less extreme ones.
- using parsimonious probability models for the propensity scores.

these reduce the variability of the probability weights at the cost of increasing the bias of the weighted estimators that use them.

Optimal inverse probability weights

Let us consider two n-dimensional vectors of probability weights, $\hat{\mathbf{w}}^* = (\hat{w}_1^*, \ldots, \hat{w}_n^*)^T$ and $\hat{\mathbf{w}} = (\hat{w}_1, \ldots, \hat{w}_n)^T$. We assume that the weighted estimator that uses $\hat{\mathbf{w}}^*$ is unbiased for Δ but has a large standard error, while the weighted estimator that uses $\hat{\mathbf{w}}$ may be biased but has a small standard error. We define $\mathbf{s} = (s_1, \ldots, s_n)^T$, with $s_i = t_i y_i$ if $t_i = 1$ and $s_i = (1 - t_i) y_i$ if $t_i = 0$, for $i = 1, \ldots, n$. The bias of $\hat{\Delta}$ with weights $\hat{\mathbf{w}}$ is

$$\mathbb{E}\left[\mathbf{s}^T \hat{\mathbf{w}} - \Delta\right] = \mathbb{E}\left[\mathbf{s}^T \hat{\mathbf{w}} - \mathbf{s}^T \hat{\mathbf{w}}^*\right] + \underbrace{\mathbb{E}\left[\mathbf{s}^T \hat{\mathbf{w}}^* - \Delta\right]}_{=0} = \mathbb{E}\left[\mathbf{s}^T (\hat{\mathbf{w}} - \hat{\mathbf{w}}^*)\right] \tag{1}$$

Equation (1) shows that minimizing the distance between $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}^*$ is equivalent to minimizing the bias of $\hat{\Delta}$ that uses $\hat{\mathbf{w}}$. This suggests obtaining $\hat{\mathbf{w}}$ by solving the following constrained optimization problem

$$\begin{aligned} \underset{\mathbf{w} \in \mathbb{R}^n}{\text{minimize}} \quad & (\mathbf{w} - \hat{\mathbf{w}}^*)^T (\mathbf{w} - \hat{\mathbf{w}}^*) \\ \text{subject to} \quad & \hat{\sigma}_{\hat{\Delta}} \leq \xi \\ & 0 \leq \mathbf{w} \leq \epsilon \end{aligned}$$

On the choice of ξ and ϵ

- *Pre-specified level of precision*
- *Level of precision obtained by using a trimmed weights/simplified model*
- *Level of precision obtained by evaluating bias-variance trade off*
- *Lagrange multiplier*
- *Medical meaning*

We recommend setting the parameter ϵ equal to the 99.9-percentile of the distribution of the target weights. This choice is due to the fact that in small sample sizes the large-sample approximation of the standard error of the ATE, $\hat{\sigma}_{\hat{\Delta}}$, may be inadequate.

Simulations - Trimming

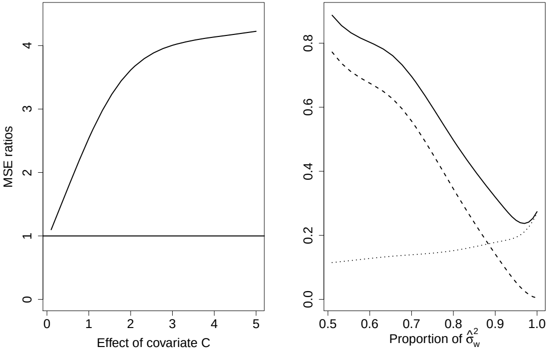


Figure : Left-hand-side panel: ratio between the observed mean squared error of the trimmed weighted mean estimator $\hat{\beta}_{2,\overline{w}}$ and that of the optimally weighted estimator $\hat{\beta}_{2,\hat{w}}$ across values of γ. Right-hand-side panel: mean squared error (solid line), variance (dotted), and bias (dashed) for different values of for different values of ξ_2 when $\gamma = 4$

Simulations - Modelling

		$n = 500$		$n = 1000$	
	Model	Optimal $\hat{\Delta}^O$	Non-optimal $\hat{\Delta}^L$	Optimal $\hat{\Delta}^O$	Non-optimal $\hat{\Delta}^L$
$\lvert E(\hat{\Delta}) - \Delta \rvert$	$\hat{\Delta}_{\hat{w}*}$	0.0557	-	0.0271	-
	$\hat{\Delta}_A$	0.0809	0.8356	0.0537	0.8326
	$\hat{\Delta}_B$	0.0590	0.2139	0.0535	0.1923
	$\hat{\Delta}_C$	0.0821	0.8892	0.0535	0.807
$\hat{\sigma}_{\hat{\Delta}}$	$\hat{\Delta}_{\hat{w}*}$	0.2919	-	0.2328	-
	$\hat{\Delta}_A$	0.2145	0.2236	0.1556	0.1600
	$\hat{\Delta}_B$	0.2535	0.2659	0.1897	0.2108
	$\hat{\Delta}_C$	0.2153	0.2250	0.1567	0.1613
95% CI coverage	$\hat{\Delta}_{\hat{w}*}$	0.9790	-	0.9810	-
	$\hat{\Delta}_A$	0.9722	0.0214	0.9647	0.0001
	$\hat{\Delta}_B$	0.9798	0.8148	0.9809	0.7515
	$\hat{\Delta}_C$	0.9724	0.0149	0.9671	0.0001

Table : Absolute bias, $|E(\hat{\Delta}) - \Delta|$, standard error, $\hat{\sigma}_{\hat{\Delta}}$, and coverage of the 95% confidence interval using the set of target weights, $\hat{\Delta}_{\hat{w}*}$, the set of optimal weights obtained by solving (2)-(2), $\hat{\Delta}^O$, and the set of non-optimal weights obtained with models A, B and C, $\hat{\Delta}^L$, with $n = 500$ and $n = 1000$

Longitudinal data

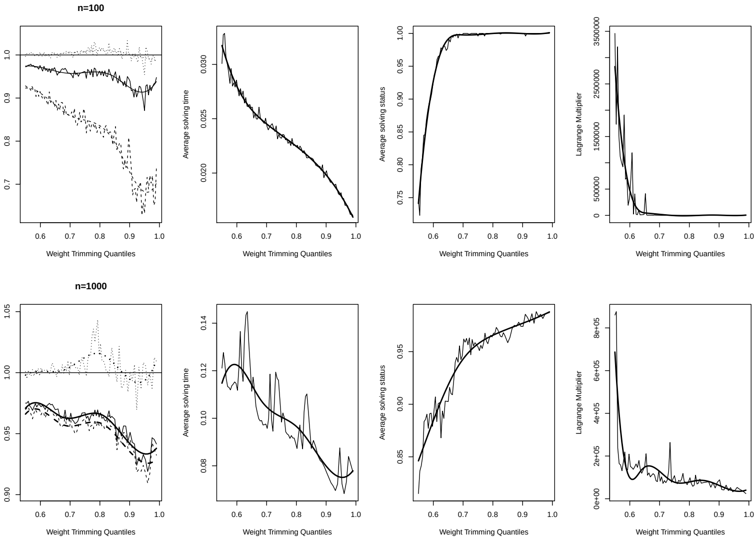
Inverse probability weighting is used with longitudinal data [6]. Suppose we observe a sample of n i.i.d. observations from a population. For each unit $i = 1, \ldots, n$, a T number of repeated measures are observed. Specifically, we observe the time-dependent variable $A_{i,t}$ and covariates $X_{i,t}$ at each time period $t = 1, \ldots, T$. Finally, we observed a sample of size n of outcome variable for all units, Y_i at the end of the study.

We can obtain $\hat{\mathbf{w}}$ by solving the following constrained optimization problem

$$\begin{aligned} \underset{\mathbf{w} \in \mathbb{R}^n}{\text{minimize}} \quad & (\mathbf{w} - \hat{\mathbf{w}}^*)^T (\mathbf{w} - \hat{\mathbf{w}}^*) \\ \text{subject to} \quad & \mathrm{Var}(\mathbf{w}) \leq \xi \\ & \mathbf{w} \geq 0 \end{aligned}$$

Similar to [5].

Simulations



Conclusions

- Probability weights are used in many settings;
- Probability weights can be highly variable;
- We propose a method to generate probability weights that are closest to target weights under specified inferential precision.
- Target weights can be estimated by using any methods for classification [2, 3, 4]
- Optimal weights can be obtaining using available R packages, such as `ipoptr`, `gurobi`, and `nloptr`
- Identification of causal effects
- More on optimal weighting coming soon [7]

References

[1] M. Santacatterina and M. Bottai, "Optimal probability weights for inference with constrained precision," *Journal of the American Statistical Association*, 2017.

[2] S. Setoguchi, S. Schneeweiss, M. A. Brookhart, R. J. Glynn, and E. F. Cook, "Evaluating uses of data mining techniques in propensity score estimation: a simulation study," *Pharmacoepidemiology and drug safety*, vol. 17, no. 6, pp. 546–555, 2008.

[3] B. K. Lee, J. Lessler, and E. A. Stuart, "Improving propensity score weighting using machine learning," *Statistics in Medicine*, vol. 29, no. 3, pp. 337–346, 2010.

[4] M. E. Karim, J. Petkau, P. Gustafson, H. Tremlett, and T. B. S. Group, "On the application of statistical learning approaches to construct inverse probability weights in marginal structural cox models: Hedging against weight-model misspecification," *Communications in Statistics - Simulation and Computation*, vol. 0, no. 0, pp. 1–30, 0.

[5] J. R. Zubizarreta, "Stable weights that balance covariates for estimation with incomplete outcome data," *Journal of the American Statistical Association*, vol. 110, no. 511, pp. 910–922, 2015.

[6] J. M. Robins, M. Ángel Hernán, and B. Brumback, "Marginal structural models and causal inference in epidemiology," *Epidemiology*, vol. 11, no. 5, pp. 550–560, 2000.

[7] N. Kallus, "Generalized optimal matching methods for causal inference," *arXiv preprint arXiv:1612.08321*, 2016.