Main Argument

1. Background material on estimating the average treatment effect from the standpoint of semiparametric models, mainly summarizing Tsiatis 2008.

    A. Introduce semiparametric estimator for the average treatment effect. Assume additional covariates are available. Introduce augmentation term, and optimized estimator that is semiparametric efficient given covariates.

    B. Discuss the "principled approach" to estimating the regression terms, by estimating $E(Y \mid A=1, X)$ and $E(Y \mid A=0, X)$ separately.

2. Show that the terms may be combined to $E(\tilde{Y} \mid X)$, eliminating the regressions on treatment. In general this requires transforming the response to $\tilde{Y}$. In case the treatment propensity is $1/2$, no transformation is required.

3. Advantages of this simplification

    A. Eliminate need for multiple teams/analysts etc under Tsiatis protocol.

    B. Better estimate the conditional expectation by simultaneously using the data from both arms

        i. Maybe a toy example here

            a. improvement in standard error from estimating $E(Y \mid X)$ directly rather than combining estimates $E(Y \mid A=0, X)$ and $E(Y \mid A=1, X)$

            b. improvement in point estimate in case X isn't well balanced over the arms

4. Implication for studies that use residuals as response or similar types of standardization. In some fields it is common to regress out a variable of interest Y on covariates, studying instead the residuals $Y' = Y - E(Y \mid X)$. The average treatment effect of Y' is obtained as the slope in the model $E(Y') = \beta_0 + \beta_1 A$.

    A. Example from epi literature

    B. Argue that the OLS estimate of β_1 is asymptotically equivalent to to the optimized estimator in case $p = P(A=1) = 1/2$.

    C. When $p \neq 1/2$, the OLS estimate is not equivalent to the optimized estimator. Since the optimized estimator is semiparametric efficient, the OLS estimate is not.

        i. The difference in asymptotic variances is

        $$(1-2p)^2 p(1-p) \operatorname{Var}(E(Y \mid A=1, X) - E(Y \mid A=0, X))$$

        which is 0 iff $p = 1/2$ or the "stratified ATEs" $E(Y \mid A = 1, X) - E(Y \mid A=0, X)$ are constant

5. We can estimate both arms at the same time for other estimators, besides the ATE. Here the advantage is in the improved efficiency of estimating both rather than combining two estimates.

   A. log-linear

   B. logistic

   C. discrete hazard?

Simulation

1. plot improved efficiency of using the semiparametric estimator over the residual OLS estimator $\hat{\beta}_1$ for a range of values $p \neq 1/2$ and values $\mathrm{Var}(E(Y \mid A = 1, X) - E(Y \mid A = 0, X) - \psi)$. The data is generated from a linear model with normal errors.

Discussion

1. some drawbacks

   A. An unscrupulous analyst may still select the model $E(Y \mid X)$ that optimizes the magnitude of the ATE estimate $\hat{psi}$ if he has access to the treatment indicators. So it may still be necessary to have two analysts, one without access who estimates $E(Y \mid X)$, and one with access who then uses the first analyst's estimated CE to compute $\hat{psi}$.

      i. There are a few small criticisms of Tsiatis's approach in the literature as well, that we might include. Eg it may be obvious to an analyst, based on the response data, who received treatment and who didn't.

   B. for the other estimators (eg logistic) consistent estimators must be used in the weights, so the improved efficiency of simultaneously estimating both arms must be balanced against the loss of (finite-sample) efficiency of plugging in the consistent estimator.