Deductive Derivation and Computerization of Semiparametric Efficient Estimation

Constantine Frangakis, Tianchen Qian,
Zhenke Wu, and Ivan Diaz

Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

JOHNS HOPKINS
BLOOMBERG SCHOOL
of PUBLIC HEALTH

November 10, 2014

Motivation

- Researchers often seek robust inference for a parameter using semiparametric estimation
- Example: estimating population mean $\mu = \mathbb{E}[Y]$ with self-selective treatments; other model parts unspecified
- A much-discussed estimator: doubly-robust estimator (Y: outcome; X: covariates; R: binary treatment indicator)
  1. Consistency: if $[R \mid X]$, or $[Y \mid R, X]$ is correct
  2. Locally efficient: has minimum possible asymptotic variance if both $[R \mid X]$, and $[Y \mid R, X]$ are correct.

Motivation

Take a step back: "What is the process to arrive at such a doubly robust estimate?"

1. Needs to know the **closed form** of estimating equations. Usually use the **efficient influence function (EIF)** for minimal asymptotic variance,
2. Identify parts of the model specification that are required in EIF. For example, the propensity score and outcome regression model.
3. Specify and fit the "working models" as guided by step 2
4. Solve for the zeros of the empirical version of EIF, after plugging in the fitted working models.
5. Hope the working models are close to truth (there are robust estimators in the presence of this deviations, e.g., Cao, Tsiatis, and Davidian, 2009)

Motivation

QUESTION:
What if, in Step 1, the EIF is hard to derive mathematically? This can happen if one is to estimate population median, or other quantiles, or other general wild estimand.

Can we specify the model and leave the efficient estimation to computer, like Stan or WinBUGS that takes in model specification, and outputs inferential results, there the posterior?

Two-stage sampling:

- Collect a set of covariates X_i, $i = 1, ..., n$
- Based on X_i, choose a subset of the first-stage subjects to measure their outcomes Y_i.
- $R_i = 1$ for those sampled in the second stage; 0 for those not included in the second stage.
- Observed data: $(X_i, Y_i R_i, R_i)$, $i = 1, ..., n$, iid
- Commonly adopted if Y_i is very expensive or invasive to measure

Goal: Estimate $\mathbb{E}[Y]$ in the presence of covariate-dependent selection into the second-stage samples

Example continued

1. We *know* the mathematical form of EIF, ϕ, is:

$$\phi\{D_i, (F-\tau), \tau\} = \frac{R_i \cdot \{Y_i - y(X_i)\}}{e(X_i)} + y(X_i) - \tau, \qquad (1)$$

2. Here we directly sove for τ following Steps 2-4:

$$\hat{\tau}^{\text{nondeductive}} = \frac{1}{n}\sum_i \frac{R_i \cdot \{Y_i - y_w(X_i)\}}{e_w(X_i)} + y_w(X_i); \qquad (2)$$

3. Our motivating question embodied in this problem:
What if (1) is unknown to us?

A deductive and compatible (DC) efficient estimation algorithm

- "Deductive": use a discrete and finite set of instructions, and, for every input, finish in discrete finite steps. (Turing, 1937)
- "Compatible": ensure estimates lie in correct range, e.g. $\mathbb{E}[Y] = Pr(Y = 1)$ will always be estimated within $[0, 1]$ if Y is binary
- "Efficient": smallest possible asymptotic variance

Inspired by

- EIF for unrestricted problem (e.g. mean outcome estimation without other model specifications) can be written in general as a Gateaux derivative (Hampel, 1974)
- "Reduction to the absurd". Use compatibility constraints to overcome the major problem of numerical computing of EIF that masks EIF's functional dependency on the target parameter.

Details of the DC-estimation procedure

Step 1 Extend the working distribution F_w to a parametric model, say, $F_w(\delta)$, around F_w (i.e., so that $F_w(0) = F_w$), where δ is a finite dimensional vector.

In practice, leave those reliable parts unmodifed, e.g., propensity score elicited by physicians.

Details of the DC-estimation procedure

Step 2 Calculate Gateaux numerical difference derivative Use the Gateaux numerical difference derivative

$$\text{Gateaux}\{\tau, F_w(\delta), D_i, \epsilon\} := \left[\tau\{F_{w(D_i,\epsilon)}(\delta)\} - \tau\{F_w(\delta)\}\right]/\epsilon$$

for a machine-small ϵ, to deduce the value of $\phi\{D_i, F_w(\delta)\}$ for arbitrary δ, and find

$$\hat{\delta}^{opt} \text{ that minimizes the empirical variance of } \tau\{F_w(\hat{\delta})\} \quad (3)$$

among all roots $\{\hat{\delta}\}$ that are subject to the condition

$$\sum_i \left[\phi\{D_i, F_w(\hat{\delta})\} \longleftarrow \text{Gateaux}\{\tau, F_w(\hat{\delta}), D_i, \epsilon\}\right] = 0, \quad (4)$$

Details of the DC-estimation procedure

Step 3 Calculate the parameter at the EIF-fitted distribution $F_w(\hat{\delta})$ as

$$\hat{\tau}^{\text{deductive}} := \tau\{F_w(\hat{\delta}^{opt})\}. \tag{5}$$

Empirical variance estimated by jacknife.

Properties

1. Consistent as the usual, non-deductive estimators (e.g. Scharfstein et al., 1999), if a variation independent component of F is specified correctly (i.e., locally semiparametrically efficient)
2. Asymptotically normal under regularity conditions
3. Does not need functional form of efficient influence function, ϕ, but deduce it by numerical Gateaux derivative
4. Extends working model and performs empirical minimization (similar to Chaffee and van der Laan (2011); Rubin and van der Laan (2008)).

Feasibility: the two-phase design problem again

Asthma care study (Huang et al, 2005)

- Mailed survey of patients from 20 California physician groups between July 1998 and Feburary 1999 to assess physician group performance
- Different distributions of characteristics across groups
- Outcome (Y): patient satisfaction for asthma care (binary, yes/no)
- Covariates (X): age, gender, race, education, health insurance, drug insurance coverage, asthma severity, number of comorbidities, and SF=36 physical and mental scores.

Goal: Compare rates of patient satisfaction for asthma care, accounting for selective behaviors into physician groups

Asthma care study continued

- To estimate the probability of patient satisfaction adjusted for covariates for two physician group pairs.
- We compare the following three estimates:
  1. unadjusted analysis
  2. adjusted analysis, non-deductive estimation (this talk)
  3. adjusted adjusted, deductive estimation

Asthma care study continued

Comparisons

estimates of $\tau(F) = \int_{x \in A} y(x)p(x)dx$

(i) all patients — $A : \{\text{all } x\}$

physician group (g)	n	unadjusted % $pr(Y = 1 \mid G = g)$	$\hat{\tau}^{\text{nondeductive}}$ (%)	se	$\hat{\tau}^{\text{deductive}}$ (%)	se
a_1	171	62.0	63.1	4.5	63.3	4.4
b_1	81	58.0	52.0	8.8	51.9	8.9
a_2	104	78.8	72.1	8.2	71.6	8.0
b_2	189	47.6	49.4	4.5	49.4	4.4

(ii) patients with increased common support — $A :$ patients with $\hat{e}(x) \in (0.1, 0.9)^{(1)}$

physician group (g)	n	unadjusted % $pr(Y = 1 \mid G = g)$	$\hat{\tau}^{\text{nondeductive}}$ (%)	se	$\hat{\tau}^{\text{deductive}}$ (%)	se
a_1	107	65.4	65.3	5.2	65.4	5.2
b_1	76	59.2	59.3	6.9	59.1	6.8
a_2	95	77.9	75.6	6.2	75.3	6.2
b_2	154	46.8	46.2	5.1	46.3	5.0

$^{(1)}$This estimand with increased "common support" (e.g., Crump et al. (2009)), excludes here 64, 5, 9 and 35 patients from a_1, b_1, a_2, b_2 respectively.

Implications

- Usual doubly-robust estimator can be re-expressed compatibly by the set of parameter values derived by the deductive estimator
- Increased "common support" gives further evidence that deductive estimator gives similar estimates as from non-deductive estimators
- No form of efficient influence function necessary, and can be computed in < 1s

Extension beyond mean outcome

Example: median for continuous outcomes

- We can as well calculate numerical Gateaux derivative
- Following the DC-estimation procedure, gives locally semiparametric efficient estimator
- No other implemented semiparametrically efficient estimator
- Simulation studies shows consistency
- Feasible for general estimand

Main points once again

1. Proposed a deductive and compatible (DC) procedure to give locally semiparametrically efficient estimates
2. Works for general estimand beyond population mean
3. Relies on numerical methods for differentiation and root finding
4. Save dramatic amounts of human efforts on essentially computerizable processes, and minimize errors.

Discussions

Extentions to restricted problems:

- Here, in the model specification, we did not constraint the form of how mean outcome depends on covariates; only working models were used as intermediate tools
- What if $E[Y_i|X_i] = \boldsymbol{\beta}' X_i$ is true and other model parts are unspecified? Our goal is to infer $\boldsymbol{\beta}$
- When Xs are discrete, we have shown that efficiency bound of $\boldsymbol{\beta}$ can be achieved by
  1. Using proposed method for unrestricted problem, and
  2. Imposing the linear restrictions numerically

Discussions

Computation:

- standard root finding methods may be unstable: use targeted MLE to transform root finding problem to iterative loss minimization problem.

Thank you!