Statistics and Probability Letters 97 (2015) 185–191

Contents lists available at ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Optimal restricted estimation for more efficient longitudinal causal inference

Edward H. Kennedy [a,*], Marshall M. Joffe [a], Dylan S. Small [b]

[a] *Department of Biostatistics and Epidemiology, University of Pennsylvania, United States*
[b] *Department of Statistics, The Wharton School, University of Pennsylvania, United States*

ARTICLE INFO

Article history:
Received 12 October 2014
Received in revised form 26 November 2014
Accepted 28 November 2014
Available online 5 December 2014

Keywords:
Doubly robust
Generalized method of moments
Marginal structural model
Semiparametric efficiency
Structural nested model
Time-varying confounding

ABSTRACT

Efficient semiparametric estimation of longitudinal causal effects is often analytically or computationally intractable. We propose a novel restricted estimation approach for increasing efficiency, which can be used with other techniques, is straightforward to implement, and requires no additional modeling assumptions.

1. Introduction

Locally efficient semiparametric estimation of causal effects in longitudinal studies can be analytically or computationally intractable; however, more simple and straightforward estimation techniques can be very imprecise. In this work we develop an approach for deriving more efficient estimators of parameters in such settings based on the idea of optimal restricted estimation, i.e., finding estimators that are optimally efficient among all those within some restricted class. In essence our approach amounts to finding optimal linear combinations of estimating functions, using constant coefficient matrices. The proposed approach can be used in conjunction with other techniques (such as those based on local efficiency derivations), is straightforward to implement, requires neither extra modeling assumptions nor extra model fitting, and comes with guarantees of better (or at least no worse) asymptotic efficiency. It can be viewed as a way to give analysts extra chances at attaining the semiparametric efficiency bound. We explore finite sample properties of our approach using simulated data.

2. Setup

Many important models in longitudinal causal inference, including structural nested models (Robins, 1989, 1994) and marginal structural models (Robins, 2000; Hernán et al., 2002), lead to estimators that solve (at least up to asymptotic

* Correspondence to: Room 507, Blockley Hall, 423 Guardian Drive, Philadelphia, PA 19104, United States.
E-mail address: kennedye@mail.med.upenn.edu (E.H. Kennedy).

http://dx.doi.org/10.1016/j.spl.2014.11.022

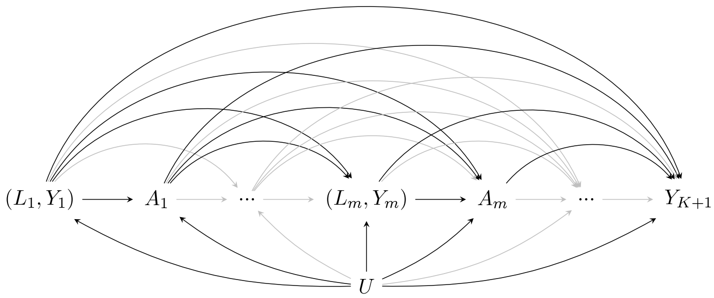

Fig. 1. Directed acyclic graph of data structure assuming only time ordering.

equivalence) estimating equations of the form

$$\mathbb{P}_n\left\{\sum_{t=1}^{K} m_t(\psi; \hat{\eta}, h)\right\} = 0$$

where $\mathbb{P}_n$ is the empirical measure so that $\mathbb{P}_n(W) = n^{-1}\sum_i W_i$ denotes a usual sample average, m_t is an estimating function of the same dimension as the parameter of interest $\psi \in \mathbb{R}^q$, η is a nuisance function taking values in some metric space, and h is an arbitrary function that affects the efficiency but not consistency of the estimator.

For example, in many settings the observed data consist of sequences of time-varying measurements of covariates L, treatment A, and outcome Y for each of n subjects. Let an overbar denote the past history of a variable so that $\overline{W}_t = (W_1, W_2, \ldots, W_t)$, and let $X_t = (\overline{L}_t, \overline{Y}_t, \overline{A}_{t-1})$ represent the observed data available just prior to treatment at time t. Also for simplicity assume no censoring and discrete measurement times $t = 1, \ldots, K$. Then a standard longitudinal study would yield an independent and identically distributed sample of observations $(Z_1, \ldots, Z_n)$, with $Z = (\overline{L}_K, \overline{A}_K, \overline{Y}_{K+1})$. Fig. 1 shows a directed acyclic graph illustrating this data structure, allowing for the presence of unmeasured variables U and only incorporating the assumed time ordering.

Let $Y_{t+1}^{\bar{a}_t}$ denote the potential outcome that would have been observed for a particular subject had that subject taken treatment sequence $\bar{a}_t$ up to time t. Then a standard repeated measures marginal structural mean model (MSMM) (Robins, 1989, 1994) assumes

$$E\left(Y_{t+1}^{\bar{a}_t} \mid V = v\right) = g_t(\bar{a}_t, v; \psi)$$

for $t = 1, \ldots, K$ and g_t specified functions known up to the parameter of interest ψ, where $V \subseteq L_1$ is an arbitrary subset of baseline covariates whose modification of the effect of treatment is of particular interest. Similarly a standard structural nested mean model (SNMM) (Robins, 2000; Hernán et al., 2002) assumes that

$$E\left(Y_{K+1}^{\bar{a}_t,0} - Y_{K+1}^{\bar{a}_{t-1},0} \mid X_t = x_t, A_t = a_t\right) = \gamma_t(x_t, a_t; \psi)$$

for $t = 1, \ldots, K$, where the specified functions γ_t (also known up to ψ) are restricted so that $\gamma_t(x_t, 0; \psi) = 0$ since $Y_{K+1}^{\bar{a}_t,0} - Y_{K+1}^{\bar{a}_{t-1},0} = 0$ if $a_t = 0$. We consider linear SNMMs for effects on the last outcome for ease of notation, but one could similarly use a log link or repeated measures models for effects on all outcomes. One could also consider versions of the above models that contrast functionals other than the mean (e.g., percentiles).

As discussed by van der Laan and Robins (2003), Tsiatis (2006), and others, under standard 'no unmeasured confounding' identifying assumptions (e.g., sequential ignorability, or $Y_{t+s}^{\bar{a}_K} \perp\!\!\!\perp A_t \mid X_t$ for $t = 1, \ldots, K$ and $s = 1, \ldots, K+1-t$), estimating functions m_t under the above MSMMs and SNMMs are given by $m_t(\psi; \eta, h) = \phi_t(\psi; \eta_a, h) - E\{\phi_t(\psi; \eta_a, h) \mid X_t, A_t\} + E\{\phi_t(\psi; \eta_a, h) \mid X_t\}$ where

$$\phi_t(\psi; \eta_a, h) = \begin{cases} h_t(\overline{A}_t, V)\left\{\dfrac{Y_{t+1} - g_t(\overline{A}_t, V; \psi)}{\prod_{s=1}^{t} p(A_s \mid X_s)}\right\} & \text{for MSMMs} \\ \left\{h_t(X_t, A_t) - \displaystyle\int h_t(X_t, a_t) p(a_t \mid X_t)\, d\nu(a_t)\right\}\left\{Y_{K+1} - \displaystyle\sum_{s=t}^{K} \gamma_s(X_s, A_s; \psi)\right\} & \text{for SNMMs,} \end{cases}$$

with the functions $p(a_t \mid x_t)$ denoting the conditional density of treatment given observed history, and ν a dominating measure for the distribution of treatment. In this setting the nuisance function $\eta = (\eta_a, \eta_y)$ consists of two variation independent components; η_a denotes the conditional treatment densities $p(a_t \mid x_t)$ and η_y denotes the conditional outcome/covariate densities $p(l_t, y_t \mid x_{t-1}, a_{t-1})$. Importantly, the functions $h_t : D_t \to \mathbb{R}^q$ (where $D_t = (\overline{A}_t, V)$ for MSMMs and $D_t = (X_t, A_t)$