Higher Order Estimating Equations for Causal Inference

Lingling Li, James Robins, Eric Tchetgen, Aad van der Vaart

http://www.math.vu.nl/∼aad

Harvard University and Vrije Universiteit Amsterdam

Higher Order Estimating Equations for Causal Inference

Lingling Li, James Robins, Eric Tchetgen, Aad van der Vaart

http://www.math.vu.nl/∼aad

Harvard University and Vrije Universiteit Amsterdam

general purpose

Introduce new type of estimating equations
to cope with high-dimensional covariates

general purpose

Introduce new type of estimating equations
to cope with high-dimensional covariates

OUTLINE

1. Introduction
2. Linear estimation equations
3. Higher order estimation equations
4. Concluding remarks

1. INTRODUCTION

example: missing data

Outcome $Y \in \{0, 1\}$
Treatment indicator $A \in \{0, 1\}$
Covariate Z
Observe (YA, A, Z)

example: missing data

Outcome $Y \in \{0,1\}$
Treatment indicator $A \in \{0,1\}$
Covariate Z
Observe (YA, A, Z)

$Y \perp\!\!\!\perp A \mid Z$
$Y \mid Z \sim$ binomial $\big(1, b(Z)\big)$
$A \mid Z \sim$ binomial $\big(1, p(Z)\big)$
$Z \sim$ density f

example: missing data

Outcome $Y \in \{0,1\}$
Treatment indicator $A \in \{0,1\}$
Covariate Z
Observe (YA, A, Z)

$Y \perp\!\!\!\perp A | \, Z$
$Y | \, Z \sim$ binomial $\big(1, b(Z)\big)$
$A | \, Z \sim$ binomial $\big(1, p(Z)\big)$
$Z \sim$ density f

Parameter (b, p, f)
Parameter of interest $\int bf = \mathrm{E}Y$ marginal treatment effect

example: missing data

Outcome $Y \in \{0, 1\}$
Treatment indicator $A \in \{0, 1\}$
Covariate Z
Observe (YA, A, Z)

$Y \perp\!\!\!\perp A \,|\, Z$
$Y \,|\, Z \sim$ binomial $\big(1, b(Z)\big)$
$A \,|\, Z \sim$ binomial $\big(1, p(Z)\big)$
$Z \sim$ density f

Parameter (b, p, f)
Parameter of interest $\int bf = \mathrm{E}Y$ marginal treatment effect

Our Interest: • Z is high-dimensional
• p, b, f are unknown

estimating equations

Data $X_1, \ldots, X_n$
Parameter of interest θ
Nuisance parameter γ
Estimator $\hat{\gamma}$

estimating equations

Data $X_1, \ldots, X_n$
Parameter of interest θ
Nuisance parameter γ
Estimator $\hat{\gamma}$

ORDINARY (= LINEAR) $$\sum_{i=1}^{n} \psi_{\theta,\hat{\gamma}}(X_i) = 0$$

estimating equations

Data $X_1, \ldots, X_n$
Parameter of interest θ
Nuisance parameter γ
Estimator $\hat{\gamma}$

ORDINARY (= LINEAR) $$\sum_{i=1}^{n} \psi_{\theta,\hat{\gamma}}(X_i) = 0$$

SECOND ORDER $$\sum\sum_{1 \leq i \neq j \leq n} \psi_{\theta,\hat{\gamma}}(X_i, X_j) = 0$$

estimating equations

Data $X_1, \ldots, X_n$
Parameter of interest θ
Nuisance parameter γ
Estimator $\hat{\gamma}$

ORDINARY (= LINEAR) $$\sum_{i=1}^{n} \psi_{\theta,\hat{\gamma}}(X_i) = 0$$

HIGHER ORDER $$\sum \cdots \sum_{1 \le i_1 \ne \cdots \ne i_k \le n} \psi_{\theta,\hat{\gamma}}(X_{i_1}, \ldots, X_{i_k}) = 0$$

estimating equations

Data $X_1, \ldots, X_n$
Parameter of interest θ
Nuisance parameter γ
Estimator $\hat{\gamma}$

ORDINARY (= LINEAR) $$\sum_{i=1}^{n} \psi_{\theta,\hat{\gamma}}(X_i) = 0$$

HIGHER ORDER $$\sum \cdots \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} \psi_{\theta,\hat{\gamma}}(X_{i_1}, \ldots, X_{i_k}) = 0$$

GENERAL FORM $$\mathbb{U}_n \psi_{\theta,\hat{\gamma}} = 0$$

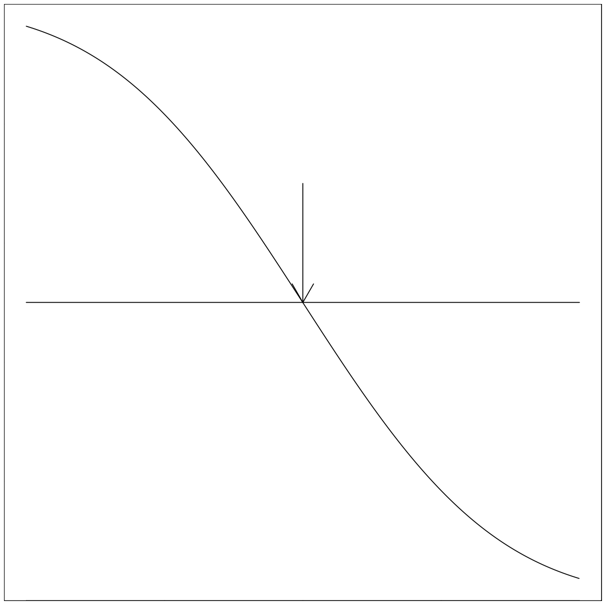
$$\mathbb{U}_n f = \frac{(n-k)!}{n!} \sum \cdots \sum_{1 \leq i_1 \neq \cdots \neq i_k \leq n} f(X_{i_1}, \ldots, X_{i_k})$$

U-statistic

one-step estimator

Next iterate in Newton scheme for finding zero
(Fisher scoring)

$$\theta \mapsto \mathbb{U}_n \psi_{\theta,\hat{\gamma}}$$

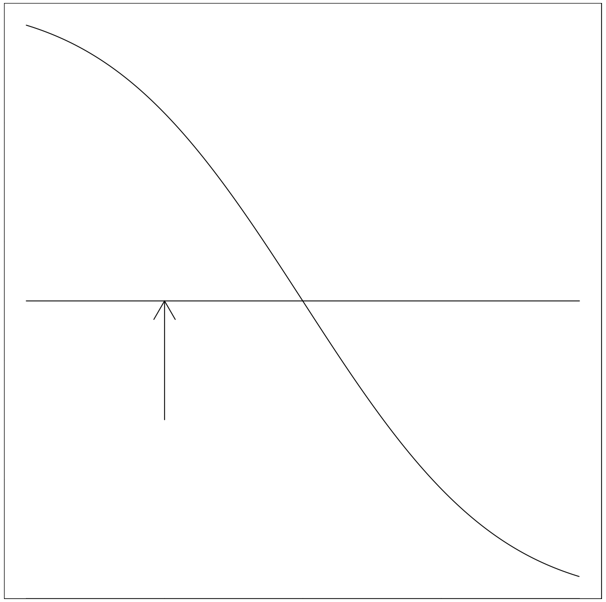


true zero

one-step estimator

Next iterate in Newton scheme for finding zero
(Fisher scoring)

$\theta \mapsto \mathbb{U}_n \psi_{\theta,\hat{\gamma}}$

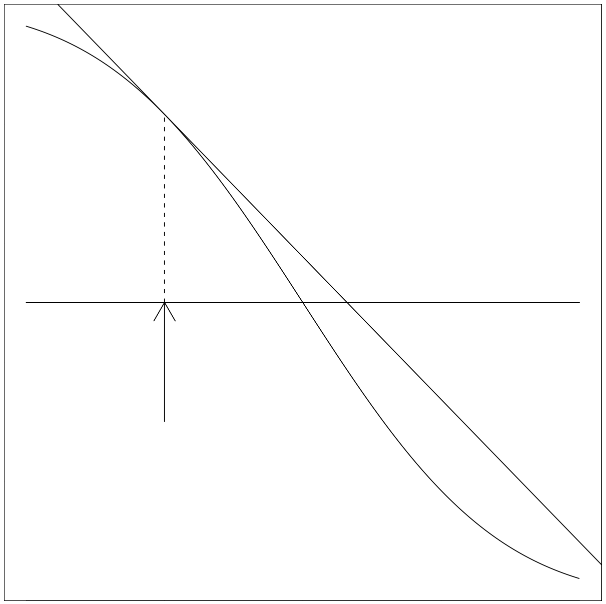


initial estimator

one-step estimator

Next iterate in Newton scheme for finding zero (Fisher scoring)

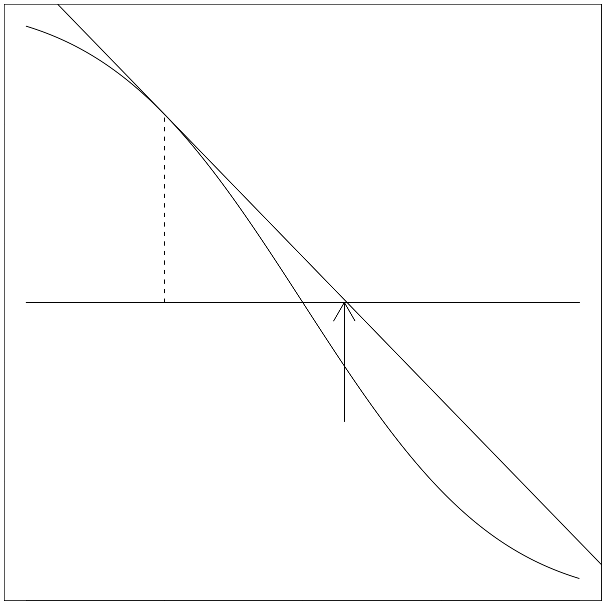
$$\theta \mapsto \mathbb{U}_n \psi_{\theta,\hat{\gamma}}$$



one-step estimator

Next iterate in Newton scheme for finding zero
(Fisher scoring)

$$\theta \mapsto \mathbb{U}_n \psi_{\theta,\hat{\gamma}}$$



one-step estimator

$$\hat{\theta} + \hat{\Sigma}_n^{-1} \mathbb{U}_n \psi_{\hat{\theta},\hat{\gamma}}$$

influence function

WANT

Influence function that works with general purpose $(\hat{\theta}, \hat{\gamma})$

influence function

WANT

Influence function that works with general purpose $(\hat{\theta}, \hat{\gamma})$

GENERAL NOTATION

Parameter $\eta = (\theta, \gamma)$

Initial estimator $\hat{\eta} = (\hat{\theta}, \hat{\gamma})$

Parameter of interest $\theta = \chi(\eta)$

Estimator $\chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$

Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

influence function

WANT

Influence function that works with general purpose $(\hat{\theta}, \hat{\gamma})$

GENERAL NOTATION

Parameter $\eta = (\theta, \gamma)$

Initial estimator $\hat{\eta} = (\hat{\theta}, \hat{\gamma})$

Parameter of interest $\theta = \chi(\eta)$

Estimator $\chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$

Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

Good influence functions have “correct” inner products (covariances) with score functions

Linear case: tangent space (semiparametrics, 1980s/90s)

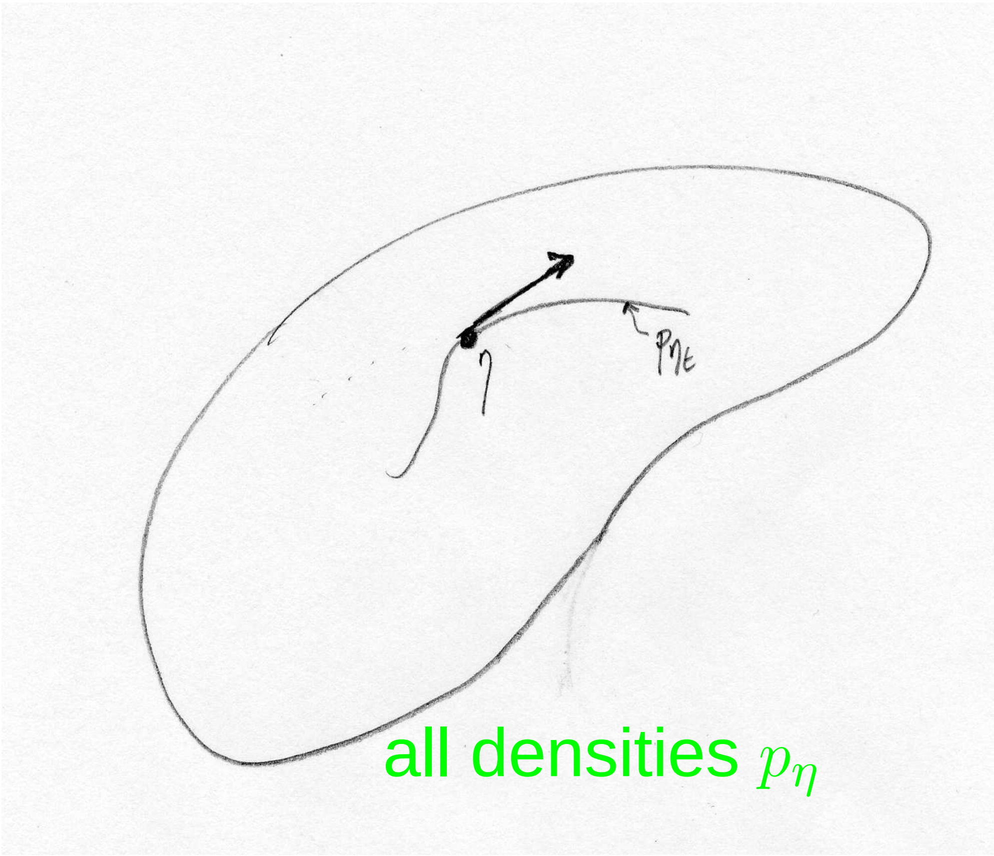
Higher-order case: higher order tangent space

2. LINEAR ESTIMATING EQUATIONS

(review)

tangent space

Tangent space (at η): all score functions $g = \frac{d}{dt}_{|t=0} \log p_{\eta_t}$ of one-dimensional submodels $t \mapsto p_{\eta_t}$



example: missing data

Outcome $Y \in \{0,1\}$
Treatment indicator $A \in \{0,1\}$
Covariate Z
Observe (YA, A, Z)

$Y \perp\!\!\!\perp A | Z$
$Y | Z \sim$ binomial $\big(1, b(Z)\big)$
$A | Z \sim$ binomial $\big(1, p(Z)\big)$
$Z \sim$ density f

Parameter $\eta = (b, p, f)$

example: missing data

Outcome $Y \in \{0, 1\}$
Treatment indicator $A \in \{0, 1\}$
Covariate Z
Observe (YA, A, Z)

$Y \perp\!\!\!\perp A \,|\, Z$
$Y \,|\, Z \sim$ binomial $\big(1, b(Z)\big)$
$A \,|\, Z \sim$ binomial $\big(1, p(Z)\big)$
$Z \sim$ density f

Parameter $\eta = (b, p, f)$

Likelihood $f(Z)p(Z)^A\big(1 - p(Z)\big)^{1-A} b(Z)^{YA}\big(1 - b(Z)\big)^{(1-Y)A}$

example: missing data

Observe (YA, A, Z)
Parameter $\eta = (b, p, f)$

Likelihood $f(Z)p(Z)^A\big(1 - p(Z)\big)^{1-A} b(Z)^{YA}\big(1 - b(Z)\big)^{(1-Y)A}$

$$\frac{A - p(Z)}{p(Z)(1-p)(Z)}\pi(Z)$$ p-score, $p_t = p + t\pi$

$$\frac{A(Y - b(Z))}{b(Z)(1-b)(Z)}\beta(Z)$$ b-score, $b_t = b + t\beta$

$$\phi(Z)$$ f-score, $f_t = f(1 + t\phi)$

influence function

Tangent space (at η): all score functions $g = \frac{d}{dt}_{|t=0} \log p_{\eta_t}$ of one-dimensional submodels $t \mapsto p_{\eta_t}$

Influence function of $\eta \mapsto \chi(\eta)$ is map $x \mapsto \dot\chi_\eta(x)$ with for all submodels $t \mapsto p_{\eta_t}$

$$\frac{d}{dt}\chi(\eta_t)_{|t=0} = \mathrm{E}_\eta g(X_1)\dot\chi_\eta(X_1)$$

influence function

Tangent space (at η): all score functions $g = \frac{d}{dt}_{|t=0} \log p_{\eta_t}$ of one-dimensional submodels $t \mapsto p_{\eta_t}$

Influence function of $\eta \mapsto \chi(\eta)$ is map $x \mapsto \dot\chi_\eta(x)$ with for all submodels $t \mapsto p_{\eta_t}$

$$\frac{d}{dt}\chi(\eta_t)_{|t=0} = \mathrm{E}_\eta g(X_1)\dot\chi_\eta(X_1)$$

- Influence function is unique up to orthocomplement of tangent space
- Minimum variance obtained with influence function inside tangent space

example: semiparametrics

Parameter $\eta = (\theta, \gamma)$

θ-Score $\dot\ell_{\theta,\gamma} = \frac{\partial}{\partial\theta} \log p_{\theta,\gamma}$

Efficient θ-score $\tilde\ell_{\theta,\gamma} = \dot\ell_{\theta,\gamma} - \Pi_{\theta,\gamma}\dot\ell_{\theta,\gamma}$

($\Pi_{\theta,\gamma}$ the orthogonal projection on the space of scores for γ)

Influence function for θ is $\tilde I_{\theta,\gamma}^{-1}\tilde\ell_{\theta,\gamma}$

(Can add orthocomplement of tangent space)

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$

Likelihood $f(Z)p(Z)^A\big(1 - p(Z)\big)^{1-A} b(Z)^{YA}\big(1 - b(Z)\big)^{(1-Y)A}$

$$\frac{A - p(Z)}{p(Z)(1-p)(Z)}\pi(Z) \qquad p\text{-score, } p_t = p + t\pi$$

$$\frac{A(Y - b(Z))}{b(Z)(1-b)(Z)}\beta(Z) \qquad b\text{-score, } b_t = b + t\beta$$

$$\phi(Z) \qquad f\text{-score, } f_t = f(1 + t\phi)$$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$

Likelihood $f(Z)p(Z)^A\big(1-p(Z)\big)^{1-A} b(Z)^{YA}\big(1-b(Z)\big)^{(1-Y)A}$

$\dfrac{A-p(Z)}{p(Z)(1-p)(Z)}\pi(Z)$ p-score, $p_t = p + t\pi$

$\dfrac{A(Y-b(Z))}{b(Z)(1-b)(Z)}\beta(Z)$ b-score, $b_t = b + t\beta$

$\phi(Z)$ f-score, $f_t = f(1+t\phi)$

Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

Influence function $\dot\chi_\eta(X) = \dfrac{A(Y-b(Z))}{p(Z)} + b(Z) - \chi(\eta)$

one-step estimator

Tangent space (at η): all score functions $g = \frac{d}{dt}_{|t=0} \log p_{\eta_t}$ of one-dimensional submodels $t \mapsto p_{\eta_t}$

Influence function of $\eta \mapsto \chi(\eta)$ is map $x \mapsto \dot\chi_\eta(x)$ with for all submodels $t \mapsto p_{\eta_t}$

$$\frac{d}{dt}\chi(\eta_t)_{|t=0} = \mathrm{E}_\eta g(X_1)\dot\chi_\eta(X_1)$$

One-step estimator $\hat\theta = \chi(\hat\eta) + \mathbb{U}_n \dot\chi_{\hat\eta}$ satisfies

$$\begin{aligned}\hat\theta - \chi(\eta) &= (\mathbb{U}_n - \mathrm{E}_\eta)\dot\chi_{\hat\eta} + \bigl[\chi(\hat\eta) - \chi(\eta) - (\mathrm{E}_{\hat\eta} - \mathrm{E}_\eta)\dot\chi_{\hat\eta}(X_1)\bigr] \\ &= O_P\Bigl(\frac{1}{\sqrt{n}}\Bigr) + O_P\bigl(\|\hat\eta - \eta\|^2\bigr)\end{aligned}$$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

One-step estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$ satisfies

$$\hat{\theta} - \chi(\eta) = (\mathbb{U}_n - \mathrm{E}_\eta)\dot{\chi}_{\hat{\eta}} + \left[\int \Big(\frac{p}{\hat{p}} - 1\Big)(\hat{b} - b) f\right]$$

$$= O_P\Big(\frac{1}{\sqrt{n}}\Big) + O_P\big(\|\hat{p} - p\| \|\hat{b} - b\|\big)$$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

One-step estimator $\hat\theta = \chi(\hat\eta) + \mathbb{U}_n \dot\chi_{\hat\eta}$ satisfies

$$\hat\theta - \chi(\eta) = (\mathbb{U}_n - \mathrm{E}_\eta)\dot\chi_{\hat\eta} + \Big[\int \Big(\frac{p}{\hat p} - 1\Big)(\hat b - b) f\Big]$$

$$= O_P\Big(\frac{1}{\sqrt{n}}\Big) + O_P\big(\|\hat p - p\| \|\hat b - b\|\big)$$

b	p	f	nobiasif
$O_P(n^{-1/2})$	$o_P(1)$	$-$	$\dim(b) < \infty$
$o_P(1)$	$O_P(n^{-1/2})$	$-$	$\dim(p) < \infty$
$n^{-\alpha/(2\alpha+d)}$	$n^{-\alpha/(2\alpha+d)}$	$-$	$\alpha > d/2$

conclusion (linear)

If using linear estimating equation with high-dimensional covariate:

- Bias dominates
- Rate $\hat{\theta} - \theta \gg n^{-1/2}$
- Linear equation suboptimal
- Confidence interval based on zero-mean normality fails

conclusion (linear)

If using linear estimating equation with high-dimensional covariate:

- Bias dominates
- Rate $\hat{\theta} - \theta \gg n^{-1/2}$
- Linear equation suboptimal
- Confidence interval based on zero-mean normality fails

REMEDY

Higher order estimating equation

3. HIGHER ORDER ESTIMATING EQUATIONS

one-step estimator

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

Good influence functions have “correct” inner products with (higher order) score functions

tangent space

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

Good influence functions have “correct” inner products with (higher order) score functions

Tangent space of order k: all higher order score functions

$$\frac{\frac{d^j}{dt^j}_{|t=0} \prod_{i=1}^k p_{\eta_t}(x_i)}{\prod_{i=1}^k p_\eta(x_i)} \qquad j = 1, \ldots, k, \quad \text{any path } t \mapsto \eta_t$$

influence function

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

Good influence functions have "correct" inner products with (higher order) score functions

Influence function should satisfy

$$\frac{d^j}{dt^j}_{|t=0} \chi(\eta_t) = \frac{d^j}{dt^j}_{|t=0} \mathrm{E}_{\eta_t} \dot{\chi}_\eta \qquad j = 1, \ldots, k, \text{ any path } t \mapsto \eta_t$$

one-step estimator

Parameter η
Initial estimator $\hat\eta$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat\theta = \chi(\hat\eta) + \mathbb{U}_n \dot\chi_{\hat\eta}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot\chi_\eta(x_1, \ldots, x_k)$

Good influence functions have “correct” inner products with (higher order) score functions

Influence function should satisfy

$$\frac{d^j}{dt^j}_{|t=0} \chi(\eta_t) = \frac{d^j}{dt^j}_{|t=0} \mathrm{E}_{\eta_t} \dot\chi_h \qquad j = 1, \ldots, k, \text{ any path } t \mapsto \eta_t$$

Expect $\hat\theta - \theta = (\mathbb{U}_n - \mathrm{E}_\eta)\dot\chi_{\hat\eta} + O_P\big(\|\hat\eta - \eta\|^{k+1}\big)$

influence function-computation

Parameter η
Initial estimator $\hat\eta$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat\theta = \chi(\hat\eta) + \mathbb{U}_n \dot\chi_{\hat\eta}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot\chi_\eta(x_1, \ldots, x_k)$

Hoeffding decomposition $\dot\chi_\eta = \dot\chi_\eta^{(1)} + \frac{1}{2}\dot\chi_\eta^{(2)} + \cdots + \frac{1}{k!}\dot\chi_\eta^{(k)}$

$\dot\chi_\eta^{(1)}$ is first order influence function of $\eta \mapsto \chi(\eta)$

$x_j \mapsto \dot\chi_\eta^{(j)}(x_1, \ldots, x_j)$ is first order influence function of
$\eta \mapsto \dot\chi_\eta^{(j-1)}(x_1, \ldots, x_{j-1})$ $\quad (j = 2, \ldots, k)$

(Optimal version may need projection in tangent space)

influence function

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

influence function

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

EXACT ONES MAY NOT EXIST

If they did, we would have bias reduction without increasing variance

influence function

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

EXACT ONES EXIST ON SUBMODELS OF $\dim < \infty$

Approximate ideal $\dot{\chi}_\eta$ by $\dot{\chi}_{\eta,m}$

influence function

Parameter η
Initial estimator $\hat{\eta}$
Parameter of interest $\theta = \chi(\eta)$
Estimator $\hat{\theta} = \chi(\hat{\eta}) + \mathbb{U}_n \dot{\chi}_{\hat{\eta}}$
Influence function $(x_1, \ldots, x_k) \mapsto \dot{\chi}_\eta(x_1, \ldots, x_k)$

EXACT ONES EXIST ON SUBMODELS OF $\dim < \infty$

Approximate ideal $\dot{\chi}_\eta$ by $\dot{\chi}_{\eta,m}$

Optimal k and submodel determined by trade-off between three terms

$$\hat{\theta} - \theta = (\mathbb{U}_n - \mathrm{E}_\eta)\dot{\chi}_{\hat{\eta},m} + O_P\left(\|\hat{\eta} - \eta\|^{k+1}\right) + \text{approximation bias}_m$$

$O_P((\mathbb{U}_n - \mathrm{E}_\eta)\dot{\chi}_{\hat{\eta},m}) \gg 1/\sqrt{n}$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

$$\dot\chi^{(1)}_{b,p,f}(X) = \frac{A(Y - b(Z))}{p(Z)} + b(Z) - \chi(\eta)$$

$$\dot\chi^{(2)}_{b,p,f}(X_1, X_2) = \Big[-\frac{A_1(Y_1 - b(Z_1))}{p(Z_1)}(A_2 - p(Z_2)) - \frac{(A_1 - p(Z_1))}{p(Z_1)}\frac{A_2(Y_2 - b(Z_2))}{p(Z_2)}\Big] K^f_m(Z_1, Z_2)$$

K^f_m projection kernel on $L_2(f)$

projection kernel

$L \subset L_2(f)$, $\quad \dim(L) = m < \infty$

$K_m : L_2(f) \to L$, projection $K_m l = l$ if $l \in L$, $K_m l = 0$ if $l \perp L$

$K_m h(x) = \int h(y) K_m(x, y)\, f(y)\, dy$

e.g. $K_m(x, y) = \sum_{i=1}^m e_i(x) e_i(y)$, $\quad e_1, e_2, \ldots$ orthonormal basis

K_m works as Dirac kernel on L: $K_m h = h$ iff $h \in L$ iff

$h(x) = \int h(y) K_m(x, y)\, f(y)\, dy$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

$$\dot\chi^{(1)}_{b,p,f}(X) = \frac{A(Y - b(Z))}{p(Z)} + b(Z) - \chi(\eta)$$

$$\dot\chi^{(2)}_{b,p,f}(X_1, X_2) = \Big[-\frac{A_1(Y_1 - b(Z_1))}{p(Z_1)}(A_2 - p(Z_2)) - \frac{(A_1 - p(Z_1))}{p(Z_1)}\frac{A_2(Y_2 - b(Z_2))}{p(Z_2)}\Big] K^f_m(Z_1, Z_2)$$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

$$\dot\chi^{(1)}_{b,p,f}(X) = \frac{A(Y - b(Z))}{p(Z)} + b(Z) - \chi(\eta)$$

$$\dot\chi^{(2)}_{b,p,f}(X_1, X_2) = \Big[-\frac{A_1(Y_1 - b(Z_1))}{p(Z_1)}(A_2 - p(Z_2)) - \frac{(A_1 - p(Z_1))}{p(Z_1)}\frac{A_2(Y_2 - b(Z_2))}{p(Z_2)}\Big] K^f_m(Z_1, Z_2)$$

VARIANCE $O(m/n^2 \vee 1/n)$
BIAS $O\big(\|\hat b - b\|\|\hat p - p\|\|\hat f - f\|\big)$
APPROXIMATION BIAS $O\big(\|b - \pi_m b\|\|p - \pi_m p\|\big)$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

WHEN USING $\dot{\chi}_{\eta}^{(k)}$

VARIANCE $O(m^{k-1}/n^k \vee 1/n)$
BIAS $O\big(\|\hat{b} - b\| \, \|\hat{p} - p\| \, \|\hat{p} - p + \hat{f} - f\|^{k-1}\big)$
APPROXIMATION BIAS $O\big(\|b - \pi_m b\| \, \|p - \pi_m p\|\big)$

example: missing data

Observe $X = (YA, A, Z)$
Parameter $\eta = (b, p, f)$
Parameter of interest $\chi(\eta) = \int bf = \mathrm{E}Y$

WHEN USING $\dot{\chi}_{\eta}^{(k)}$

VARIANCE $O(m^{k-1}/n^k \vee 1/n)$
BIAS $O\left(\|\hat{b} - b\| \|\hat{p} - p\| \|\hat{p} - p + \hat{f} - f\|^{k-1}\right)$
APPROXIMATION BIAS $O\left(\|b - \pi_m b\| \|p - \pi_m p\|\right)$

α	m	k
≥ 5	n	1
$[2.5, 5)$	n	2
$[\cdots, 2.5)$	$n^{15\alpha/(2\alpha+10)}$	3

$\dim(Z) = 10$

confidence intervals

Confidence intervals can be based on (conditional) asymptotic normality of $\hat{\theta} - \theta$

confidence intervals-monte carlo

regularity		length*100			coverage			truncation	
p,b	f	$k=1$	$k=2$	$k=3$	$k=1$	$k=2$	$k=3$	$k=2$	$k=3$
6	2	2.64	3.09	3.60	89	91	88	1	1
5	2	2.64	3.18	3.34	85	90	90	1	1
3	6	2.56	3.41	3.81	39	88	93	1	1
3	4	2.97	3.34	3.36	43	89	94	1	1
3	2	2.95	3.30	3.45	51	89	91	1	1
2.5	4	3.29	3.72	3.75	13	83	87	1	1
2.5	2	2.91	3.39	3.34	16	86	91	1.05	1
1.25	4	3.11	9.08	7.56	0	82	77	1.33	1.20
1.25	2	3.23	8.69	7.40	0	84	76	1.33	1.20
3.125	0.625	2.50	4.08	4.24	53	86	85	1.12	1

Regularity relative to $\dim(Z) = 10$, Nominal coverage 90 %

Sample size 1000, Kernel based on Daubechies 6

4. CONCLUDING REMARKS

adaptation

If true parameter η cannot be assumed regular, consistent estimation of $\chi(\eta)$ is impossible

Rate of precision depends on assumed regularity level
(higher precision with higher regularity)

adaptation

If true parameter η cannot be assumed regular, consistent estimation of $\chi(\eta)$ is impossible

Rate of precision depends on assumed regularity level
(higher precision with higher regularity)

ESTIMATION Adaptation to regularity possible
(single procedure gives best rate for range of regularity levels, e.g. cross validation)

CONFIDENCE INTERVALS Adaptation severely limited
(length of interval depends on a lowest level of regularity)

MORE

Talks by Lingling Li and Eric Tchetgen

10.50, room 103A