

General Formulations of Semiparametric Missing-Data & Calibrated Problems

GENERAL OBJECTIVE: formulate iid problems with ‘survey-like’ features, and define what it means for a survey estimator to be large-sample optimal.

iid data $(X_i^{(1)}, R_i, R_i(Y_i, X_i^{(2)})), 1 \leq i \leq n$

These may be observed data or a **biased sample** using a known function $w(X_i^{(1)})$ so that underlying data are drawn not from $f_{X,R,Y}((x^{(1)}, x^{(2)}), r, y)$ but from

$$f_{X,R,Y}(x, r, y) w(x^{(1)}) / \iiint f_{X,R,Y}(u, t, v) w(u^{(1)}) du dt dv$$

For discussion of estimation from biased sampling models:

Gill, R., Vardi, Y. and Wellner (1988, Ann. Statist.)
projecteuclid.org/euclid.aos/1176350948.

In the data triple, the **covariate vector** $X_i^{(1)}$ is always observed, $R_i = 0, 1$ is a **response indicator**, and Y_i is the **attribute of interest**.

Target of estimation is $E(Y_1)$.

APPROACH VIA PARAMETRIC MODELS, LIKELIHOOD

Specify joint density of the data through

$$X \sim f_0(x, \eta_1) \text{ wrt } \tau(dx)$$

$$Y \text{ given } X \sim f(y - \mu_1 - m_1(x) | x, \eta_2) \text{ wrt } \nu(d\epsilon)$$

$$\epsilon \equiv Y - \mu_1 - m_1(X)$$

$$P(R = 1 | X, Y) = P(R = 1 | X) \equiv \pi(X, \eta_3)$$

SUBJECT TO

$$\int m_1(x) p_0(x) d\tau(x) = 0$$

$$\int \epsilon f(\epsilon | x) d\nu(\epsilon) = 0$$

AND MAYBE (as in Chen & Qin 1993)

$$\int w(x) g(x) d\tau(x) = 0$$

RECALL FROM CHAPTER 3 (PARAMETRIC)
FROM TSIATIS (2006) BOOK

Regard $\mu_1 = E(Y)$ as scalar parameter of interest,
and parameterize *nuisance* functions

$$m_1(x, \beta), f_0(x, \eta_1), f(\epsilon|x; \eta_2), \pi(x, \eta_3)$$

If $X_i = X_i^{(1)}$, then likelihood is $L(\underline{X}, \underline{R}, \underline{RY}) =$

$$\prod_{i=1}^n [f_0(X_i) \pi(X_i)^{R_i} (1-\pi(X_i))^{1-R_i} f(Y_i - \mu_1 - m_1(X_i) | X_i)^{R_i}]$$

subject to constraints. RAL estimators must satisfy

$$\tilde{\mu}_1 - \mu_1 = \frac{1}{n} \sum_{i=1}^n \varphi(X_i, R_i, R_i Y_i) + o_P(1/\sqrt{n})$$

for **influence function** $\varphi(\cdot)$ such that

$$\varphi(X, R, RY) \perp \Lambda_{\beta, \eta_1, \eta_2, \eta_3} = \text{cls}(S_\beta, S_{\eta_1}, S_{\eta_2}, S_{\eta_3})$$

and

$$\varphi(X, R, RY) = \Pi\left(-R \frac{f'(\epsilon|X)}{f(\epsilon|X)} \mid \Lambda_{\beta, \eta_1, \eta_2, \eta_3}^\perp\right)$$

plus a component orthogonal to all of the parametric scores.

This leads (Robins and Rotnitzky 1995, Z. Tan 2006, 2007, Tsiatis 2006) to optimal estimator of the form

$$\varphi = \frac{R}{\pi(X)} (Y - m_1(X) - \mu_1) + m_1(X)$$

and suggests (borne out in other Z. Tan papers) that with the additional calibration constraints $E(w(X)) = 0$, the optimal estimators will combine:

mean-centered outcome predictions $\hat{m}_1(X_i)$

inverse prob-weighted resid's $\frac{R_i}{\hat{\pi}(X_i)} (Y_i - \hat{\mu}_1 - \hat{m}_1(X_i))$

calibration residuals $\frac{R_i}{\hat{\pi}(X_i)} (X_i - \hat{X})$

ADDITIONAL DETAILS IN MODIFIED FORMULATION

Now forget about μ_1 and let the parameter of interest be β in the **Outcome Model** $Y = \mu(X, \beta) + \epsilon$, but retain distinction between $X^{(1)}, X^{(2)}$.

NB. $Y, \epsilon \in \mathbf{R}^d, \beta \in \mathbf{R}^q$

Reference: Chapters 4 and 7–9 of Tsiatis (2006).

Consider **influence functions** $\varphi(X_i^{(1)}, R_i, R_i(Y_i, X_i^{(2)}))$

\propto functions spanned by scores $\frac{\partial}{\partial \theta_j} \log L$

\perp *nuisance tangent space* spanned by $S_{\eta_1}, S_{\eta_2}, S_{\eta_3}$.

(I) *Restricted moment model:* with $\pi(x, \eta_3) \equiv 1$,

(Thm 4.8) $\Lambda_{\eta_1, \eta_2}^\perp = \{A_{q \times d}(X)(Y - \mu(X, \beta)) : \text{any } A\}$

influence fcns $[E(A(X)D(X))]^{-1} A(X) (Y - \mu(X, \beta_0))$

where

$$D(X) = \frac{\partial \mu(X, \beta_0)}{\partial \beta^T} = J_{\mu(X, \cdot) | \beta_0}$$

Unique efficient influence function uses

$$A(X) = D(X)^T [E(\{Y - \mu(X, \beta_0)\}^{\otimes 2} | X)]^{-1}$$

(II) *Restricted moment with nonresponse by design:* η_3 known, $\pi(x, \eta_3) > 0$,

(Thm 7.2) influence fcns $[E(A(X)D(X))]^{-1} \cdot$

$$\left[\frac{R}{\pi(Y, X^{(1)})} A(X) (Y - \mu(X, \beta_0)) + \left(\frac{R_i}{\pi(Y, X^{(1)})} - 1 \right) L(X^{(1)}) \right]$$

(III) *Models for Outcome and Nonresponse:* η_3 unknown, $\pi(x, \eta_3) > 0$,

(Thm 8.3) influence fcns

$$\begin{aligned} &\propto \left[\frac{R}{\pi(Y, X^{(1)})} A(X) (Y - \mu(X, \beta_0)) \right. \\ &\quad \left. + Q(X, Y)R + Q^*(X^{(1)}) (1 - R) \right] \\ &\quad - \Pi([\cdot] | \text{span}(S_{\eta_3})) \end{aligned}$$

where

$$S_{\eta_3} = \frac{\partial \pi}{\partial \eta_3} \cdot \frac{R - \pi}{\pi(1 - \pi)}$$

IDEA OF ESTIMATING EQUATIONS WITH SPECIFIED NUISANCE FUNCTIONS

(A) Taylor linearization about β_0 says that the solution of the estimating equation

$$\begin{aligned} \sum_{i=1}^n \left[\frac{R_i}{\pi(Y_i, X_i^{(1)})} A(X_i) (Y_i - \mu(X_i, \beta)) + \right. \\ \left. + \left(\frac{R_i}{\pi(Y_i, X_i^{(1)})} - 1 \right) L(X_i^{(1)}) \right] = 0 \end{aligned}$$

has influence function as in **(II)** above (cases of non-response by design)

(B) Also: if $\hat{\eta}_3$ is estimated via MLE in $\text{Binom}(1, \pi(Y_i, X_i^{(1)}))$ conditional likelihood for R_i , then get influence function of form in **(III)** above for

$$\begin{aligned} \sum_{i=1}^n \left[\frac{R_i}{\pi(Y_i, X_i^{(1)}, \hat{\eta}_3)} A(X_i) (Y_i - \mu(X_i, \beta)) + \right. \\ \left. + Q(X_i, Y_i, X_i) R_i + Q^*(X_i^{(1)}) (1 - R_i) \right] = 0 \end{aligned}$$

ALTERNATIVE: ADJUSTED-WEIGHT CALIBRATION

$$\min_w \sum_{i=1}^n R_i d_i G_i(w_i, d_i)$$

subject to

$$\frac{1}{n} \sum_{i=1}^n R_i w_i X_i = \frac{1}{n} t_{\mathbf{X}} \quad \text{or} \quad E(X)$$

followed by estimator $n^{-1} \sum_{i=1}^n R_i w_i Y_i$.

Empirical likelihood takes $G_i(w_i, d_i) = -\log(w_i/d_i)$,

and has the connection with *iid* sampling that the distribution of data triples is approximated by a discrete distribution with masses p_i closely related to w_i . The idea is that (apart from additive constant) the empirical likelihood approximates $-\sum_{i \in \mathcal{U}} \log w_i$ and weights w_i are naturally normalized so that $\sum_{i=1}^n w_i \equiv N$.

FURTHER COMMENTS ON INFLUENCE FUNCTIONS AND ESTIMATING EQUATIONS

(1) Always there is only one ‘best’ influence function in the sense of having minimum variance for the influence functions of RAL Estimators

$$\hat{\mu}_1 \quad \text{or} \quad \hat{\beta} = \frac{1}{n} \sum_{i=1}^n \varphi(Z_i) + o_P\left(\frac{1}{\sqrt{n}}\right)$$

EFFICIENT INFLUENCE FUNCTION always has form

$$\varphi(Z) = \{E(S_{eff}(Z)S_{eff}(Z)^{tr})\}^{-1} S_{eff}(Z)$$

where

$$S_{eff}(Z) = S_{\mu_1} - \Pi[S_{\mu_1} | \Lambda_{nuis}] = \Pi[S_{\mu_1} | \Lambda_{nuis}^\perp]$$

Yet Tsiatis’ book is filled with results about characterizing the (unnormalized) influence function space Λ_{nuis}^\perp .

This is because the efficient score always involves true but unknown parameters for which we plug in estimators. This changes the direction of the influence function, but generally gives another influence function [not necessarily the efficient one!] for another RAL estimator.

2) ‘Paradox’ of reduced efficiency in using true parameter values in estimating equation

Paper of Henmi and Eguchi Biometrika 2004) isolated the problem as follows:

Theorem 1: assume that the estimating function $u(z, \theta)$ for parameters $\theta = (\beta, \alpha)$ and possible nuisance parameters κ has the property that

$$\Pi(s_\theta(z) | u(z, \theta)) \equiv (u_{*\beta}(x, \theta), u_{*\alpha}(x, \theta))$$

with orthogonal components $u_{*\beta}, u_{*\alpha}$. Then with $\tilde{\beta}$ denoting the estimating-equation estimator with fixed α and $\hat{\beta}$ the usual estimating-equation estimator,

$$avar(\tilde{\beta}) \geq avar(\hat{\beta})$$

generally with inequality !!

This happens typically if the likelihood factors $L(\theta, \kappa) = L_1(\beta, \kappa) \cdot L_2(\alpha)$ and we take $u_\alpha(z, \theta) \equiv s_\alpha(x, \theta)$ which is what happens with our missing-data propensities under MAR.

(3) Ryan Janicki (2009 UMCP PhD) thesis covers cases of submodels within (semi-)parametric models estimated via estimating equations.

He shows (in current work being written into paper with A. Kagan) that in the parametric estimating equation context, there is a way to generate optimal estimating equation for submodel in which this phenomenon of (2) cannot happen !

Notation: estimating function $u(z, \theta)$

with $B_u = E(u(Z) u(Z)^{tr})$

and $C_u = -E(\partial u / \partial \theta)$ nonsingular

Projection $u_*(z, \theta) = C_u^{tr} B_u^{-1} u$

In submodel $\theta = \theta(\eta)$ with $m = \dim(\eta) < \dim(\theta) = s$

the optimal estimating function is:

$$D^{tr} C_u^{tr} B_u^{-1} u \quad \text{for} \quad D^{s \times m} = \frac{\partial \theta}{\partial \eta}$$

So if we want to fix components α and let $\eta = \beta$, $\theta = (\beta, \alpha_0)$, then in general the optimal estimating equation does not involve just using the β portion of the original estimating function !

Likelihood vs Estimating Equation Methods

Tsiatis (pp. 160-163, Ch.7) considers the cases where there is some missing data, say in the form $(R_i X_i, R_i, Y_i)$ with response propensity model $\pi(X_i, \alpha)$ and outcome model $p(y|x, \beta)$ likelihood may involve integrals so that direct maximization in a semiparametric or even high-dimensional setting may be impossible.

Survey setting:

Consider case of $(R_i, X_i^{(1)}, R_i X_i^{(2)}, R_i Y_i)$ where response-propensity models properly depend on both $X^{(1)}, X^{(2)}$ but only $X^{(1)}$ is seen in advance. Maybe we can assume

$$p(x^{(2)} | x^{(1)}, r = 1) = p(x^{(2)} | x^{(1)})$$

as well as the MAR-like assumption

$$p(y | x^{(1)}, x^{(2)}, r = 1) = p(y | x)$$

Then propensity and outcome models may be identifiable, but likelihood methods are unlikely to work well.