Semiparametric Adaptive Estimation With Nonignorable Nonresponse Data

[Survey Group Working Seminar@Iowa State University]

Kosuke Morikawa\textsuperscript{1}    Jae Kwang Kim\textsuperscript{2}

March 31, 2017

\textsuperscript{1}Graduate School of Engineering Science, Osaka University, Japan

\textsuperscript{2}Department of Statistics, Iowa State University, USA and Department of Mathematical Sciences, KAIST, Korea
Estimation Under Nonignorable Nonresponse Data

• $Y$: response variable and $X$: covariate vector

• We want to estimate $E(Y)$ under nonignorable nonresponse

• Classical approaches require correct specification of $[Y \mid X]$ as well as that of the response mechanism
  • This assumption on $[Y \mid X]$ is to be relaxed or eliminated

• We propose two types of semiparametric estimators which does not demand the assumption and also attain the semiparametric efficiency bound.
Outline

Introduction

Semiparametric Estimation

Proposed Method

Simulation

Summary
Introduction
What are Nonignorable Nonresponse Data?

- **Nonresponse Data**
  - $Y$: response variable (missing)
  - $X$: covariate vector
  - $R$: response indicator of $Y$

- **Nonignorable Data (NMAR)**
  - Response mechanism $P(R = 1 \mid x, y)$ depends on $Y$
  - If the mechanism does NOT depend on $Y$, it is called ignorable or MAR
When $\pi$ is known

- If the response mechanism $\pi(x, y) = P(R = 1 \mid x, y)$ is known, we can easily estimate $E(Y)$ (Ex.) IPW estimator

$$n^{-1} \sum_{i=1}^{n} \frac{r_i y_i}{\pi(x_i, y_i)} \xrightarrow{p} E\left( \frac{R Y}{\pi(X, Y)} \right) = E\left( \frac{E(R \mid X, Y) Y}{\pi(X, Y)} \right) = E(Y)$$

- Generally speaking, however, the mechanism is unknown and to be estimated!
When $\pi$ is known

- If the response mechanism $\pi(x, y) = P(R = 1 \mid x, y)$ is known, we can easily estimate $E(Y)$ (Ex.) IPW estimator

\[
n^{-1} \sum_{i=1}^{n} \frac{r_i y_i}{\pi(x_i, y_i)} \xrightarrow{p} E \left( \frac{R Y}{\pi(X, Y)} \right) = E \left( \frac{E(R \mid X, Y) Y}{\pi(X, Y)} \right) = E(Y)
\]

- Generally speaking, however, the mechanism is unknown and to be estimated!
When $\pi$ is known

- If the response mechanism $\pi(x, y) = P(R = 1 \mid x, y)$ is known, we can easily estimate $E(Y)$ (Ex.) IPW estimator

$$n^{-1} \sum_{i=1}^{n} \frac{r_i y_i}{\pi(x_i, y_i)} \xrightarrow{p} E \left( \frac{R Y}{\pi(X, Y)} \right) = E \left( \frac{E(R \mid X, Y) Y}{\pi(X, Y)} \right) = E(Y)$$

- Generally speaking, however, the mechanism is unknown and to be estimated!
Maximum Likelihood Method – Complete ver.–

- Likelihood:
  \[ \prod_{i=1}^{n} P(R_i = 1 \mid x_i, y_i; \phi)^{r_i} P(R_i = 0 \mid x_i, y_i; \phi)^{1-r_i} \]

- Score equation:
  \[ \sum_{i=1}^{n} \left\{ r_i \frac{\partial \log \pi_i(\phi)}{\partial \phi} + (1 - r_i) \frac{\partial \log(1 - \pi_i(\phi))}{\partial \phi} \right\} = 0 \]

\[ \Leftrightarrow \sum_{i=1}^{n} \left\{ r_i s_{1,i}(\phi) + (1 - r_i) s_{0,i}(\phi) \right\} = 0, \]

where

\[ s_{r,i}(\phi) = s_{r,i}(x_i, y_i; \phi) = \frac{r_i - \pi_i(\phi)}{\pi_i(\phi) \{1 - \pi_i(\phi)\}} \]

and \(\pi_i(\phi) = \pi(x_i, y_i; \phi)\) and \(\dot{\pi} = \frac{\partial \pi(\phi)}{\partial \phi}\)
• Observed likelihood:

\[
\prod_{i=1}^{n} P(R_i = 1 \mid x_i, y_i; \phi)^{r_i} P(R_i = 0 \mid x_i; \phi)^{1-r_i}
\]

\[
= \prod_{i=1}^{n} P(R_i = 1 \mid x_i, y_i; \phi)^{r_i} \left\{ \int P(R_i = 0 \mid y, x_i; \phi) f(y \mid x_i) dy \right\}^{1-r_i}
\]

• Score equation:

\[
\sum_{i=1}^{n} \left\{ r_is_{1,i}(\phi) + (1 - r_i) \frac{-E[\hat{\pi}(x_i, Y; \phi) \mid x_i]}{E[1 - \pi(x_i, Y; \phi) \mid x_i]} \right\} = 0
\]

• Thus, if \( f(y \mid x) \) is known, \( \phi \) can be estimated, but the assumption is too strong...
Some previous (semiparametric) methods to estimate the parameter $\phi$ of a response mechanism $\pi(x, y; \phi)$:

- $f_1(y \mid x) := f(y \mid x, r = 1)$ is necessary
  - Riddles et al. (2016, Surv. Methodol.)
    - Though this method requires a stronger assumption, this is more efficient (because it is based on likelihood method)
    - Misspecification of $f_1$ leads to an inconsistent estimator of $\phi$ and $\theta$
  
- $f(y \mid x)$ is unnecessary
  - Qin et al. (2002, JASA)
  - Chang & Kott (2008, Biometrika)
• Recall that the score equation is
\[ \sum_{i=1}^{n} \left\{ r_i s_{1,i}(\phi) + (1 - r_i) \frac{-E[\hat{\pi}(x_i, Y; \phi) \mid x_i]}{E[1 - \pi(x_i, Y; \phi) \mid x_i]} \right\} = 0. \]

• For any function \( g(x, y) \), by Bayes’ formula, it holds that
\[ E(g(x, Y) \mid x) = \frac{E_1[\pi(x, Y)^{-1} g(x, Y) \mid x]}{E_1[\pi(x, Y)^{-1} \mid x]}, \]
where \( E_1(\cdot \mid x) = E(\cdot \mid x, r = 1) \).

• Thus, the second term is transformed to
\[ \frac{-E_1[\pi(x_i, Y; \phi)^{-1} \hat{\pi}(x_i, Y; \phi) \mid x_i]}{E_1[\pi(x_i, Y; \phi)^{-1} \{1 - \pi(x_i, Y; \phi)\} \mid x_i]} = \frac{E_1[O(x_i, Y; \phi) s_0(x_i, Y; \phi) \mid x_i]}{E_1[O(x_i, Y; \phi) \mid x_i]} \]
where \( O(x, y) = (1 - \pi(x, y))/\pi(x, y) \).
• Recall that the score equation is

\[
\sum_{i=1}^{n} \left\{ r_i s_{1,i} (\phi) + (1 - r_i) \frac{-E[\hat{\pi}(x_i, Y; \phi) \mid x_i]}{E[1 - \pi(x_i, Y; \phi) \mid x_i]} \right\} = 0.
\]

• For any function \(g(x, y)\), by Bayes’ formula, it holds that

\[
E(g(x, Y) \mid x) = \frac{E_1[\pi(x, Y)^{-1} g(x, Y) \mid x]}{E_1[\pi(x, Y)^{-1} \mid x]},
\]

where \(E_1(\cdot \mid x) = E(\cdot \mid x, r = 1)\).

• Thus, the second term is transformed to

\[
-\frac{E_1[\pi(x_i, Y; \phi)^{-1} \hat{\pi}(x_i, Y; \phi) \mid x_i]}{E_1[\pi(x_i, Y; \phi)^{-1} \{1 - \pi(x_i, Y; \phi)\} \mid x_i]} = \frac{E_1[O(x_i, Y; \phi) s_0(x_i, Y; \phi) \mid x_i]}{E_1[O(x_i, Y; \phi) \mid x_i]}
\]

where \(O(x, y) = (1 - \pi(x, y))/\pi(x, y)\).
• Thus, the score equation can be written as

\[ \sum_{i=1}^{n} \left\{ r_i s_{1,i}(\phi) + (1 - r_i) \frac{E_1\{O(\phi)s_{0,i}(\phi) \mid x_i\}}{E_1\{O(\phi) \mid x_i\}} \right\} = 0. \]

• Therefore, it can be computed by using \( f_1(y \mid x; \gamma) \)

• Estimated \( \hat{\phi} \) has consistency and asymptotic normality (CAN), if the model \( f_1 \) is correct
Let $\hat{\phi}$ be a solution of the following equation:

$$
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) g(x_i) = 0,
$$

where $g(x)$ is an arbitrary function of $x$ whose dimension is same as $\phi$.

Under regularity conditions, $\hat{\phi}$ has CAN.

Note that $E[\{1 - r/\pi(X, Y)\}g(X)] = 0$ for all $g$.

(Ex.) Let $\pi(y; \phi) = 1/\{1 + \exp(\phi_0 + \phi_1 y)\}$ and $X$ is completely observed. Then, $(\phi_0, \phi_1)$ can be estimated by solving

$$
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(y_i; \phi)} \right) \left( \begin{array}{c} 1 \\ x_i \end{array} \right) = 0.
$$
• Let \( \hat{\phi} \) be a solution of a following equation:

\[
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) g(x_i) = 0,
\]

where \( g(x) \) is an arbitrary function of \( x \) whose dimension is same as \( \phi \)

• Under regularity conditions, \( \hat{\phi} \) has CAN

• Note that \( E[\{1 - r/\pi(X, Y)\}g(X)] = 0 \) for all \( g \).

• (Ex.) Let \( \pi(y; \phi) = 1/\{1 + \exp(\phi_0 + \phi_1y)\} \) and \( X \) is completely observed. Then, \( (\phi_0, \phi_1) \) can be estimated by solving

\[
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(y_i; \phi)} \right) \begin{pmatrix} 1 \\ x_i \end{pmatrix} = 0.
\]
Summary of Previous Works

- Riddles et al. (2016)’s method
  - Not semiparametric estimation
  - would be more efficient than Chang & Kott (2008)’s method
  - misspecification of $f_1$ leads to inconsistency
  - Not optimal

- Chang & Kott (2008) ’s method
  - Semiparametric estimation
  - Not optimal
We propose two types of semiparametric estimators.

1. with working model $f_1$
   - $[f_1 \text{ is correct}]$ : optimal
   - $[f_1 \text{ is wrong}]$ : CAN

2. without any assumptions
   - optimal
Semiparametric Estimation
• We discuss efficiency in regular and asymptotically linear (RAL) estimators (Bickel et al., 1997; Tsiatis, 2006)
• An estimator \( \hat{\theta} \) is called asymptotically linear if there exists \( \varphi \) such that
\[
\sqrt{n}(\hat{\theta} - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi(X_i, Y_i; \theta_0) + o_p(1)
\]
• It is known that it is reasonable to restrict attention to RAL estimators by Hájek’s representation theorem (Hájek, 1970)
Example of Semiparametric Estimator

- Suppose that \((X, Y)\) are completely observed.
- GEE (Generalized Estimating Equation) gives a semiparametric efficient estimator for \(E(Y \mid X) = \mu(X; \beta)\):

\[
\sum_{i=1}^{n} D^\top(X_i)V^{-1}(X_i)(Y - \mu(X_i; \beta)) = 0,
\]

where \(D(X) = \partial \mu(X; \beta)/\partial \beta^\top\) and \(V(X) = \text{var}(Y \mid X)\).

- Generally speaking, \(V(X)\) is unknown, thus we put a working model on it.
- The above solution has CAN even if the working model \(V(X)\) is incorrect. If it is correct, it attains the semiparametric efficiency bound.
Example of Semiparametric Estimator

• Suppose that \((X, Y)\) are completely observed.

• GEE (Generalized Estimating Equation) gives a semiparametric efficient estimator for \(E(Y \mid X) = \mu(X; \beta)\):

\[
\sum_{i=1}^{n} D^\top(X_i) V^{-1}(X_i)(Y - \mu(X_i; \beta)) = 0,
\]

where \(D(X) = \partial \mu(X; \beta) / \partial \beta^\top\) and \(V(X) = \text{var}(Y \mid X)\).

• Generally speaking, \(V(X)\) is unknown, thus we put a working model on it.

• The above solution has CAN even if the working model \(V(X)\) is incorrect. If it is correct, it attains the semiparametric efficiency bound.
Example of Semiparametric Estimator

- Suppose that \((X, Y)\) are completely observed.
- GEE (Generalized Estimating Equation) gives a semiparametric efficient estimator for \(E(Y \mid X) = \mu(X; \beta)\):

\[
\sum_{i=1}^{n} D^\top(X_i)V^{-1}(X_i)(Y - \mu(X_i; \beta)) = 0,
\]

where \(D(X) = \partial \mu(X; \beta)/\partial \beta^\top\) and \(V(X) = \text{var}(Y \mid X)\).
- Generally speaking, \(V(X)\) is unknown, thus we put a working model on it.
- The above solution has CAN even if the working model \(V(X)\) is incorrect. If it is correct, it attains the semiparametric efficiency bound.
Suppose that \((X, Y)\) are completely observed.

GEE (Generalized Estimating Equation) gives a semiparametric efficient estimator for \(E(Y \mid X) = \mu(X; \beta)\):

\[
\sum_{i=1}^{n} \underbrace{D^\top(X_i)V^{-1}(X_i) (Y - \mu(X_i; \beta))}_{A(X_i): A \text{ is an arbitrary function}} = 0,
\]

where \(D(X) = \partial\mu(X; \beta)/\partial\beta^\top\) and \(V(X) = \text{var}(Y \mid X)\).

Generally speaking, \(V(X)\) is unknown, thus we put a working model on it.

The above solution has CAN even if the working model \(V(X)\) is incorrect. If it is correct, it attains the semiparametric efficiency bound.
Proposed Method
The semiparametric lower bound under nonignorable nonresponse has already been derived by Rotnitzky & Robins (1997, Stat. Med.)

- They derived a “best” estimating equation for regression parameters $E(Y \mid X) = \mu(X; \beta)$
- But it requires many working models
- Misspecified working models lead to inefficient estimator

We consider an estimator for $\theta = E\{U(X,Y)\}$, where $U$ is a known function such as $Y$ and $I(Y > 0)$.

In this talk, for simplicity, let $\theta = E(Y)$
Theorem 1. (Morikawa & Kim, 2016)

Under regularity conditions, a solution \((\hat{\phi}, \hat{\theta})\) of a following estimating equation is the semiparametric efficient estimator of \((\phi_0, \theta_0)\):

\[
\sum_{i=1}^{n} \left(1 - \frac{r_i}{\pi(x_i, y_i; \phi)}\right) g_{\text{eff}}(x_i; \phi) = 0,
\]

\[
\sum_{i=1}^{n} \left\{ \frac{r_i(\theta - y_i)}{\pi(x_i, y_i; \phi)} + \left(1 - \frac{r_i}{\pi(x_i, y_i; \phi)}\right) E^* (\theta - Y \mid x_i) \right\} = 0,
\]

where \(g_{\text{eff}}(x; \phi) = E^* (s_0(x, Y; \phi) \mid x)\) and

\[
E^* \{g(x, Y) \mid x\} = \frac{E \{O(x, Y) g(x, Y) \mid x\}}{E \{O(x, Y) \mid x\}}.
\]
• $g_{\text{eff}}$ and $E^*(Y \mid x)$ are to be estimated!

• By using a similar idea to Riddles et al. (2016):

$$g_{\text{eff}}(x; \phi) = \frac{E\{O(x, Y)s_0(x, Y) \mid x\}}{E\{O(x, Y) \mid x\}} = \frac{E_1\{\pi^{-1}(x, Y)O(x, Y)s_0(x, Y) \mid x\}}{E_1\{\pi^{-1}(x, Y)O(x, Y) \mid x\}}, \quad (*)$$

where $E_1(\cdot \mid x) = E(\cdot \mid x, r = 1)$. 
1. Assume working models for $f_1(y \mid x; \gamma)$ and estimate $\hat{\gamma}$ by ML
2. Choose one model by AIC
3. $g_{\text{eff}}(\phi, \hat{\gamma})$ is computed with $f_1(y \mid x; \hat{\gamma})$.

$$g_{\text{eff}}(\phi, \hat{\gamma}) = \frac{E_1\{\pi^{-1}(x, Y)O(x, Y)s_0(x, Y) \mid x; \hat{\gamma}\}}{E_1\{\pi^{-1}(x, Y)O(x, Y) \mid x; \hat{\gamma}\}},$$

4. $g_{\text{eff}}(\phi, \hat{\gamma})$ is numerically calculated by EM algorithm (Riddles et al., 2016)

$E^*(Y \mid x)$ can be computed in a similar way.
Theorem 2. (Morikawa & Kim, 2016)

Assume a parametric model of $f_1(y \mid x; \gamma)$ and let $\hat{\gamma}$ be its MLE. Also, let $(\hat{\phi}, \hat{\theta})$ be a solution of a following estimating equation:

$$
\frac{1}{n} \sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) g_{\text{eff}}(x_i; \phi; \hat{\gamma}) = 0,
$$

$$
\frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{r_i(\theta - y_i)}{\pi(x_i, y_i; \phi)} + \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) E^*(\theta - Y \mid x_i; \hat{\gamma}; \phi) \right\} = 0.
$$

Then, $(\hat{\phi}, \hat{\theta})$ has CAN even if $f_1(y \mid x; \gamma)$ is wrong. Especially, if the assumed model is correct, the estimator attains the semiparametric efficiency bound.

- In Riddles et al. (2016), misspecification of $f_1$ leads to an inconsistent estimator of $\theta$.
- This estimator is not doubly robust.
Theorem 2. (Morikawa & Kim, 2016)

Assume a parametric model of \( f_1(y \mid x; \gamma) \) and let \( \hat{\gamma} \) be its MLE. Also, let \((\hat{\phi}, \hat{\theta})\) be a solution of a following estimating equation:

\[
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) g_{\text{eff}}(x_i; \phi; \hat{\gamma}) = 0,
\]
\[
\sum_{i=1}^{n} \left\{ \frac{r_i (\theta - y_i)}{\pi(x_i, y_i; \phi)} + \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) E^*(\theta - Y \mid x_i; \hat{\gamma}; \phi) \right\} = 0.
\]

Then, \((\hat{\phi}, \hat{\theta})\) has CAN even if \( f_1(y \mid x; \gamma) \) is wrong. Especially, if the assumed model is correct, the estimator attains the semiparametric efficiency bound.

- In Riddles et al. (2016), misspecification of \( f_1 \) leads to an inconsistent estimator of \( \theta \).
- This estimator is not doubly robust.
Some Cases When Computation of $E^*$ is Easy

- There exists a class of functions of response mechanism and $f_1(y|x)$ that does not need the computation.
- Let $f_1$ belongs to exponential family, i.e.,

$$f_1(y | x; \tau, \psi) = \exp \left\{ \frac{y \tau - b(\tau)}{\psi} + c(y, \psi) \right\} \quad (1)$$

and its mean be $\lambda\{\mu(\tau | x; \gamma)\} = \gamma^T x$ where $\lambda$ is a link function, $b$, and $c$ are known functions.
- Let response mechanism can be written as

$$\pi(x, y; \phi) = \frac{1}{1 + \exp\{h(x; \phi_0) + \phi_1 y\}} \quad (2)$$

where $\phi = (\phi_0^T, \phi_1)^T$ and $h$ is a known function.
Some Cases When Computation of $E^*$ is Easy

- There exists a class of functions of response mechanism and $f_1(y|x)$ that does not need the computation.
- Let $f_1$ belongs to exponential family, i.e.,

$$f_1(y \mid x; \tau, \psi) = \exp \left\{ \frac{y \tau - b(\tau)}{\psi} + c(y, \psi) \right\} \quad (1)$$

and its mean be $\lambda\{\mu(\tau \mid x; \gamma)\} = \gamma^\top x$ where $\lambda$ is a link function, $b$, and $c$ are known functions.
- Let response mechanism can be written as

$$\pi(x, y; \phi) = \frac{1}{1 + \exp\{h(x; \phi_0) + \phi_1 y\}}, \quad (2)$$

where $\phi = (\phi_0^\top, \phi_1^\top)^\top$ and $h$ is a known function.
**Proposition 1.** (Morikawa & Kim, 2016)

If $f_1$ belongs to exponential family (1) and $\pi$ is written as (2), then

$$g_{\text{eff}}(x)^\top = -\frac{(\dot{h}(x; \phi_0)^\top, \dot{b}(\phi_1 \psi + \tau))}{1 + d(\phi_1, \psi, \tau)}$$

and

$$E^*(Y \mid x) = \frac{\dot{b}(\phi_1 \psi + \tau) + d(\phi_1, \psi, \tau)\dot{b}(2\phi_1 \psi + \tau)}{1 + d(\phi_1, \psi, \tau)},$$

where $\dot{b}(\tau) = \partial b(\tau)/\partial \tau$ and

$$d(\phi_1, \psi, \tau) = \exp[h(x; \phi_0) + \{b(2\phi_1 \psi + \tau) - b(\phi_1 \psi + \tau)\}/\psi].$$

- In this case, EM algorithm is unnecessary
(Ex.) $f_1: \text{Normal, } \pi: \text{Logistic}$

- $Y \mid x, r = 1 \sim N(\gamma^\top x, \sigma^2)$
- $\pi(x, y) = \{1 + \exp(\phi_0 + \phi_1 y)\}^{-1}$
- By the proposition, we have

$$g_{\text{eff}}(x) = \frac{1}{1 + d(\phi_1, \sigma^2, \gamma) \left( \phi_1 \sigma^2 + \gamma^\top x \right)}$$

and

$$E^*(Y \mid x) = \frac{(\phi_1 \sigma^2 + \gamma^\top x) + d(\phi_1, \sigma^2, \gamma)(2\phi_1 \sigma^2 + \gamma^\top x)}{1 + d(\phi_1, \sigma^2, \gamma)},$$

where $d(\phi_1, \sigma^2, \gamma) = \exp\{\phi_0 + \phi_1(3\phi_1 \sigma^2/2 + \gamma^\top x)\}$
Sketch Proof of Proposition 1

- By the assumptions on $f_1$ and $\pi$, we have

$$g_{\text{eff}}(x; \phi) = \frac{E_1\{\pi^{-1}(x, Y)O(x, Y)s_0(x, Y) \mid x\}}{E_1\{\pi^{-1}(x, Y)O(x, Y) \mid x\}}$$

$$= \frac{E_1[\exp\{h(x; \phi_0) + \phi_1 Y\}(\dot{h}(x; \phi_0)^\top, Y)^\top \mid x]}{E_1[\exp\{h(x; \phi_0) + \phi_1 Y\} + \exp\{h(x; \phi_0) + 2\phi_1 Y\} \mid x]}$$

- $E_1\{\exp(\phi_1 Y) \mid x\}$ and $E_1\{Y \exp(\phi_1 Y) \mid x\}$ are to be computed!

- These are just the moment generating function and its first derivative on $f_1$. 
When $X$ is discrete, $f_1$ is not needed to be modeled

For example, when $X$ is dichotomous,

$$E_1\{\pi^{-1}(x, Y)O(x, Y) \mid x = 0\} \approx \frac{1}{\#\{R = 1, X = 0\}} \sum_{i: r_i = 1, x_i = 0} \pi^{-1}(0, y_i)O(0, y_i)$$

(*) can be computed!!
**Continuous & Nonparametric case**

- Denominator and numerator of (\(\ast\)) can be estimated nonparametrically by Nadaraya-Watson estimator:

\[
\hat{g}_{\text{eff}}(x; \phi) = \frac{\sum_{j=1}^{n} r_j \pi^{-1}(x, y_j) O(x, y_j) s_0(x, y_j) K_h(x - x_j)}{\sum_{j=1}^{n} r_j \pi^{-1}(x, y_j) O(x, y_j) K_h(x - x_j)},
\]

where \(K_h(x) = K(x/h)\) and \(K(\cdot)\) is a kernel function and bandwidth, which satisfy regularity conditions.

- (\(\ast\)) can be computed!!
**Theorem 3. (Morikawa & Kim, 2016)**

Let \((\hat{\phi}, \hat{\theta})\) be a solution of the following estimating equation:

\[
\sum_{i=1}^{n} \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) \hat{g}_{\text{eff}}(x_i; \phi) = 0,
\]

\[
\sum_{i=1}^{n} \left\{ \frac{r_i(\theta - y_i)}{\pi(x_i, y_i; \phi)} + \left( 1 - \frac{r_i}{\pi(x_i, y_i; \phi)} \right) \hat{E}^*(\theta - Y \mid x_i; \phi, \theta) \right\} = 0,
\]

where each \(\hat{g}_{\text{eff}}(\phi)\) and \(\hat{E}^*(\phi, \theta)\) is a nonparametric estimator of \(g_{\text{eff}}(\phi)\) and \(E^*(\theta - Y \mid x)\), respectively. Then under regularity conditions, \((\hat{\phi}, \hat{\theta})\) attains the semiparametric efficiency bound.
We focus on variance estimation for the first estimator. Let the estimating equation be
\[ S_{\text{eff}}^* = (S_1(\gamma^*, \phi)^\top, S_2(\gamma^*, \phi, \theta)^\top)^\top, \]
where \( \gamma^* \) is the limit of \( \hat{\gamma} \).

The asymptotic variance in our method is given as, so called sandwich estimator,
\[
E \left( \frac{\partial S_{\text{eff}}^*}{\partial (\phi^\top, \theta)} \right)^{-1} E(S_{\text{eff}}^{*\otimes 2}) E \left( \frac{\partial S_{\text{eff}}^*}{\partial (\phi^\top, \theta)} \right)^{-1, \top}.
\]

By simple algebra, the asymptotic variance of \( \hat{\theta} \) is given by
\[
V^* = \text{var}\{S_2(\gamma^*, \phi_0) - \kappa^* S_1(\gamma^*, \phi_0, \theta_0)\},
\]
where \( \kappa^* = \kappa_1^* \kappa_2^* \), \( \kappa_1^* = E\{(Y - E^*(Y \mid x; \gamma^*)) \hat{\pi}(\phi_0)/\pi(\phi_0)\} \),
and \( \kappa_2^* = E\{g_{\text{eff}}(\gamma^*, \phi_0) \hat{\pi}(\phi_0)/\pi(\phi_0)\} \).
• Unknown values $\kappa^*_1$ and $\kappa^*_2$ are expectation of known functions.

• Thus, our method can be applied again to estimate $\kappa$ by a similar way to estimate $E(Y)$

• Therefore, the asymptotic variance of $\hat{\theta}$ can be estimated by

$$\hat{V} = \frac{1}{n} \sum_{i=1}^{n} \{ S_{2i}(\hat{\gamma}, \hat{\phi}) - \hat{\kappa} S_{1i}(\hat{\gamma}, \hat{\phi}, \hat{\theta}) \} \otimes^2,$$

where $B \otimes^2 = BB^\top$. 
Simulation
Setup

We conduct a Monte Carlo simulation to show performances of our proposed estimators:

- $X = (X_1, X_2)$
- Response mechanism is NMAR:

  $$P(R = 1 \mid x_1, y) = \frac{1}{1 + \exp(\phi_0 + \phi_1 x_1 + \phi_2 y)}$$

- Target parameter: $\phi_2$ and $\theta = E(Y)$
- $Y \mid x_1, x_2, r = 1 \sim N(\mu(x_1, x_2), 1)$
- Sample size: 500, 2000
- Iteration: 2000
• **Scenario 1.** \((X_1, X_2)\): binary, \(Y\): continuous
  - \(X_1, X_2 \sim B(1/2)\)
  - \(\mu(x) = 0.5I_{00} - 0.5I_{01} - 0.5I_{10} + I_{11}\), where
    \[I_{ij} = I(X_1 = i, X_2 = j)\]
  - \((\phi_0, \phi_1, \phi_2) = (-1, -0.5, 0.75)\)

• **Scenario 2.** \(X_1\): binary, \(X_2\): continuous, \(Y\): continuous
  - \(X_1 \sim B(1/2), X_2 \sim U(-1, 1)\)
  - \(\mu(x) = \exp(0.5 - x_1 + 1.5x_2)\)
  - \((\phi_0, \phi_1, \phi_2) = (-1.7, -0.4, 0.5)\)
Methods

Assume that the response mechanism is correctly specified except for the case of MAR

- **MAR**: $\pi(x) = \{1 + \exp(\phi_0 + \phi_1 x_1 + \phi_2 x_2)\}^{-1}$

- **CK**: Chang & Kott (2008)'s method. $g(x) = (1, x_1, x_2)^\top$

- **RKI**: Riddles et al. (2016)'s method: In scenario 2,
  
  $Y \mid x_1, x_2, r = 1 \sim N(\eta_0 + \eta_1 x_1 + \eta_2 x_2, \sigma^2)$

- **Nm**: In scenario 2,
  
  $Y \mid x_1, x_2, r = 1 \sim N(\eta_0 + \eta_1 x_1 + \eta_2 x_2, \sigma^2)$

- **N**: In scenario 2,
  
  $Y \mid x_1, x_2, r = 1 \sim N(\exp(\eta_0 + \eta_1 x_1 + \eta_2 x_2), \sigma^2)$

- **NNP**: nonparametric method
Scenario1 $[\phi_2]$
Scenario 1 \( [E(Y)] \)
Scenario $\phi_2$

Estimated Value

- $\phi_2$
- $\phi_2^{2K}$
- $\phi_1$
- $\phi_1^{2K}$
- $\phi_3$
- $\phi_3^{2K}$
- $\phi_4$
- $\phi_4^{2K}$
- $\phi_5$
- $\phi_5^{2K}$
Scenario2 \[E(Y)\]
Summary
Summary

• We proposed two types of estimators which attain the semiparametric efficiency bound
  • Although first estimator requires a model on $f_1(y \mid x; \gamma)$, even estimators with misspecified models have CAN
  • Second estimator does not need the model on $f_1(y \mid x)$

Future Works

• Extension to the framework of
  • repeated measure data
  • causal inference
Main References

Questions?