A Propensity-Score-Adjustment Method for Nonignorable Nonresponse

Junho Han

MathSci, KAIST

June 12, 2019

Table of Contents

Propensity Score Method for Nonresponse

Nonresponse has become a major problem in sample surveys as participation rates have declined in many surveys. Weighting adjustments are commonly used to adjust for unit nonresponse.

Classical approaches include poststratification (Holt and Smith 1979), regression weighting (Bethlehem 1988), and raking ratio estimation (Deville, Särndal, and Sautory 1993). **Propensity-score weighting**, which increases the sampling weights of the respondents using their inverse response probabilities, is a popular approach for handling unit nonresponse.

$$U_{ps}\left(\theta\right) = \sum_{i=1}^{n} \delta_i \pi_i^{-1} U_i\left(\theta\right) = 0.$$

Propensity Score Method for Nonresponse

Setup

- x_is are always observed and y_i is observed iff $\delta_i = 1$.
- The response mechanism is $\pi_i(\phi; x_i, y_i) = pr(\delta_i = 1 | x_i, y_i)$.

Ignorable Nonresponse (missing at random)

$$\sum_{i=1}^{n} \delta_i \pi_i^{-1}\left(\hat{\phi}; x_i\right) U_i(\theta; x_i, y_i) = 0.$$

Nonignorable Nonresponse

$$\sum_{i=1}^{n} \delta_i \pi_i^{-1}\left(\hat{\phi}; x_i, \boldsymbol{y_i}\right) U_i(\theta; x_i, y_i) = 0.$$

Existing Methods

<u>Generalized Method of Moments</u>

Get $\hat{\phi}_{ck}$ from the following calibration condition:

$$\sum_{i=1}^{n} \left\{ \delta_i \pi_i^{-1} (\phi; x_i, y_i) - 1 \right\} (1, x_i) = 0.$$

This method is efficient since it doesn't assmue any outcome models, i.e. $f(y|x;\beta)$. It is first proposed by *Chang and Kott(2008)*.

Existing Methods

Fully Parametric

Get ϕ_{mle} from the mean score function (normal approach):

$$\bar{\boldsymbol{S}}\left(\phi\right) = \sum_{i=1}^{n} \left[\delta_i \boldsymbol{s_i}\left(\phi; y_i\right) + \left(1-\delta_i\right) E\left\{\boldsymbol{s_i}\left(\phi; Y\right)\left|x_i, \delta_i = 0\right.\right\}\right].$$

To solve it, we need $f\left(y|x, \delta = 0\right)$. Fully parametric approach specifies $f\left(y|x; \beta\right)$ parametrized by β to derive

$$f\left(y|x, \delta = 0\right) = f\left(y|x\right) \frac{pr\left(\delta = 0|x, y\right)}{E\left\{pr\left(\delta = 0|x, Y\right)|x\right\}}.$$

However, this can be very sensitive to failure of the assumed model because it assumes the whole model over unobserved data.

Proposed Method

Summary

- It assumes only $f(y|x,\delta=1;\gamma)$ which is relatively easy to verify from the observed part of samples rather than $f(y|x;\beta)$. So it is more robust than the fully parametric approach for an incorrectly specified outcome model.
- It is more efficient than both GMM and FP in terms of accuracy or variance of estimates.
- It suggests a novel computational tool applied to the empirical distribution of the response mechanism.

Proposed Method

Let $f_i(y|x) = f(y|x, \delta = i)$ and $E_i\{\cdot\} = E\{\cdot|\delta = i\}$. The outcome model under nonresponse can be represented as

$$f_0(y|x) = f_1(y|x)\frac{\boldsymbol{O}(x,y)}{E_1\{\boldsymbol{O}(x,Y)|x\}}$$

where $\boldsymbol{O}(x,y) = pr(\delta = 0|x,y;\phi)/pr(\delta = 1|x,y;\phi)$. Using this formula, the mean score function can be computed by

$$\bar{\boldsymbol{S}}(\phi) = \sum_{i=1}^{n}\left[\delta_i \boldsymbol{s_i}(\phi; y_i) + (1-\delta_i)\frac{E_1\{\boldsymbol{s_i}(\phi; Y)\,\boldsymbol{O}(x, Y)\,|x_i\}}{E_1\{\boldsymbol{O}(x, Y)\,|x_i\}}\right].$$

Proposed Method

Before getting $\hat{\phi}_p$ from $\bar{\boldsymbol{S}}(\phi) = 0$, we need to compute a consistent estimator $\hat{\gamma}$ for $f_1(y|x) = f_1(y|x;\gamma)$. Indeed, $\bar{\boldsymbol{S}}(\phi) = \bar{\boldsymbol{S}}(\phi, \gamma)$.

Since γ only involves the observed part of samples, $\hat{\gamma}$ is a solution of

$$\boldsymbol{S}(\gamma) = \sum_{i=1}^{n} \delta_i \boldsymbol{s_i}(\gamma) = \sum_{i=1}^{n} \delta_i \frac{\partial \log f_1(y_i|x_i;\gamma)}{\partial \gamma} = 0.$$

This is a relatively easy step compared to the main part if f_1 is appropriately specified. After finding $\hat{\gamma}$, we let $\bar{\boldsymbol{S}}(\phi) = \bar{\boldsymbol{S}}(\phi, \hat{\gamma})$.

Proposed Method

$$\bar{\boldsymbol{S}}(\phi) = \sum_{i=1}^{n} \left[\delta_i \boldsymbol{s_i}(\phi; y_i) + (1 - \delta_i) \frac{E_1 \left\{ \boldsymbol{s_i}(\phi; Y)\, \boldsymbol{O}(x, Y) \,|x_i \right\}}{E_1 \left\{ \boldsymbol{O}(x, Y) \,|x_i \right\}} \right].$$

However, the expectation part is computationally challenging; so we use 2-steps of imporance sampling technique to resolve it.

Let $\boldsymbol{Q}(x, y) = \boldsymbol{s_i}(\phi; Y)\, \boldsymbol{O}(x, Y)$ or $\boldsymbol{O}(x, Y)$, and we approximate each of $E_1 \{ \boldsymbol{Q}(x, Y) | x_i \}$.

Proposed Method

Step 1

$$E_1\left\{\boldsymbol{Q}(x, Y)|x_i\right\} \approx n_r^{-1} \sum_{\delta_j=1} \boldsymbol{Q}\left(x_i, y_j\right) \frac{f_1\left(y_j|x_i\right)}{f_1\left(y_j\right)}$$

Step 2

$$f_1\left(y_j\right) = \int f_1\left(y_j|x\right) f_1\left(x\right) dx \approx n_r^{-1} \sum_{\delta_k=1} f_1\left(y_j|x_k\right)$$

Applying this two importance sampling approximations to the original equation, we earn new approximate mean score function $\bar{\boldsymbol{S}}_2\left(\phi\right)$. Moreover, we could estimate ϕ by EM algorithm.

Proposed Method

Algorithm

1. Get $\hat{\gamma}$ from $\boldsymbol{S}(\gamma) = \sum_{i=1}^{n} \delta_i \frac{\partial \log f_1(y_i | x_i; \gamma)}{\partial \gamma} = 0.$
2. With $\bar{\boldsymbol{S}}_2(\phi) = \bar{\boldsymbol{S}}_2(\phi, \hat{\gamma})$, obtain $\hat{\phi}_p$ by EM algorithm:
$$\hat{\phi}^{(t+1)} \leftarrow \text{ solve } \bar{\boldsymbol{S}}_2\left(\phi \middle| \hat{\phi}^{(t)}\right) = 0.$$
3. Get $\hat{\theta}_p$ by propensity-score method:
$$U_{ps}(\theta) = \sum_{i=1}^{n} \delta_i \pi_i^{-1}\left(\hat{\phi}_p\right) U_i(\theta) = 0.$$

Asymptotic Properties

Theorem 1 (Asymptotic Normality)

$$\sqrt{n}\left(\hat{\phi}_p - \phi_0\right) \to N\left(0, \Sigma_\phi\right)$$
$$\sqrt{n}\left(\hat{\theta}_p - \theta_0\right) \to N\left(0, \sigma_\theta^2\right)$$

For variance estimation, we can use a linearization method. By Theorem 1, the variance of the propensity-score estimator, $\hat{\theta}_{PS,p}$, can be estimated by

$$\hat{V}_{\text{lin}}\left(\hat{\theta}_{PS,p}\right) = n^{-1}\hat{\boldsymbol{\tau}}^{-1}\hat{V}_{Ul}(\hat{\boldsymbol{\tau}}^{-1})^{T}, \tag{20}$$

where $\hat{\boldsymbol{\tau}} = n^{-1}\sum_{i=1}^{n} \delta_i \pi^{-1}\left(\mathbf{x}_{1i}, y_i; \hat{\phi}\right)\dot{u}\left(\hat{\theta}_{PS,p}; \mathbf{x}_i, y_i\right)$, $\dot{u}(\theta; \mathbf{x}, y) = \partial u(\theta; \mathbf{x}, y)/\partial\theta^{T}$, $\hat{V}_{Ul} = (n-1)^{-1}\sum_{i=1}^{n}(\hat{u}_{li} - \bar{u}_n)^2$, $\hat{u}_{li} = u_{li}(\hat{\theta}_{PS,p}, \hat{\boldsymbol{\phi}}_p, \hat{\boldsymbol{\gamma}})$, $\bar{u}_n = n^{-1}\sum_{i=1}^{n}\hat{u}_{li}$,

$$\begin{aligned}
u_{li}(\theta, \boldsymbol{\phi}, \boldsymbol{\gamma}) &= -\hat{\mathbf{B}}\bar{s}_0^{*}(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma}) \\
&+ \delta_i\left[\frac{u(\theta; \mathbf{x}_i, y_i)}{\pi(\mathbf{x}_{1i}, y_i; \boldsymbol{\phi})} - \hat{\mathbf{B}}\{\mathbf{s}(\boldsymbol{\phi}; \delta_i, \mathbf{x}_{1i}, y_i) - \bar{\mathbf{s}}_0^{*}(\boldsymbol{\phi}; \mathbf{x}_i, \boldsymbol{\phi}, \boldsymbol{\gamma}) - \hat{\boldsymbol{\kappa}}\mathbf{s}_1(\boldsymbol{\gamma}; \mathbf{x}_i, y_i)\}\right],
\end{aligned}$$

Asymptotic Properties

The variance of the propensity-score estimator, $\hat{\theta}_{PS,p}$, *can be estimated by*

$$\hat{V}(\hat{\theta}_{PS,p}) = \hat{V}_1 + \hat{V}_2,$$

where

$$\hat{V}_1 = \sum_{i \in A} \sum_{i \in A} \Omega_{ij} \hat{\eta}_i \hat{\eta}_j,$$

$$\hat{V}_2 = \sum_{i \in A} d_i \delta_i (1 - \hat{\pi}_i) \left[u_{PS,i}\left(\hat{\theta}; \hat{\boldsymbol{\phi}}\right) + \hat{\mathbf{B}}\{\mathbf{s}_{2i}\left(\hat{\boldsymbol{\phi}}; \hat{\boldsymbol{\gamma}}\right) - \hat{\boldsymbol{\kappa}}\mathbf{s}_{1i}(\hat{\boldsymbol{\gamma}})\}\right]^2,$$

Simulation Study

Setup

- $x \sim N(0, 0.5)$ and $y = m(x) + e$ with four different mean structures and three different error models:

- $m_1(x) = -1 + x,$
  $m_2(x) = -2 + 0.5\exp(0.5 + x),$
  $m_3(x) = -1 + \sin(2x),$
  $m_4(x) = -1 + 0.4x^3,$

- $e \sim N(0, 0.9),$
  $e \sim N\left(0, 0.49\left(1 + x^2\right)\right),$
  $e \sim \ln N(-0.49/2, 0.49).$

Simulation Study

Setup

- $\delta_i \sim \text{Ber}(\pi_i)$
  where $\pi_i = \{1 + \exp(-\phi_0 - \phi_1 y_i)\}^{-1}$
  with $(\phi_0, \phi_1) = (0.8, -0.2)$

- estimate $\theta = E(Y)$ with the following five methods:
  1. Full Sample (assume all data observed)
  2. Missing At Random (ignorable nonresponse)
  3. Fully Parametric (FP)
  4. Generalized Method of Moments (GMM)
  5. New Method

Simulation Study

(1) Full: Simple mean estimator with full sample.

(2) MAR: Naive estimator under the missing-at-random (MAR) assumption.

$$\hat{\theta}_{\text{MAR}} = \frac{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_m) y_i}{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_m)},$$

where $\hat{\boldsymbol{\phi}}_m$ is the maximum likelihood estimator of $\boldsymbol{\phi} = (\phi_0, \phi_1)$ assuming ignorable nonresponse, $\text{pr}(\delta = 1 | x, y) = 1 + \exp(-\phi_0 - \phi_1 x)^{-1}$.

(3) FP: Fully parametric approach using Monte Carlo EM algorithm assuming that $y|x \sim N(\beta_0 + \beta_1 x, \sigma^2)$ and a correct response model for all cases.

(4) GMM: Use the method of Chang and Kott (2008).

$$\hat{\theta}_{\text{GMM}} = \frac{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_{ck}) y_i}{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_{ck})},$$

where $\hat{\boldsymbol{\phi}}_{ck}$ is the solution to the following calibration condition,

$$\sum_{i=1}^{n} \{\delta_i \pi_i^{-1}(\boldsymbol{\phi}) - 1\}(1, x_i) = \mathbf{0}.$$

(1) New: Proposed estimator using the maximum likelihood estimator of $\boldsymbol{\phi}$.

$$\hat{\theta}_{\text{NEW}} = \frac{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_p) y_i}{\sum_{i=1}^{n} \delta_i \pi_i^{-1}(\hat{\boldsymbol{\phi}}_p)},$$

where $\hat{\boldsymbol{\phi}}_p$ is the solution to (13) assuming that the respondents' model $y|(x, \delta = 1) \sim N(\beta_0 + \beta_1 x, \sigma^2)$for all cases.

n = 500, B = 2000

Table 1. Monte Carlo Means and Variances of the Estimators of θ under Model 1

$m(x)$	Method	Mean	Variance	$m(x)$	Method	Mean	Variance
Case 1	Full	−1.001	0.0027	Case 3	Full	−0.999	0.0029
	MAR	−1.050	0.0035		MAR	−1.059	0.0037
	FP	−1.004	0.0047		FP	−1.000	0.0073
	GMM	−1.004	0.0049		GMM	−0.998	0.0075
	New	−1.003	0.0047		New	−0.998	0.0074
Case 2	Full	−0.941	0.0033	Case 4	Full	−0.999	0.0023
	MAR	−1.003	0.0041		MAR	−1.053	0.0030
	FP	−0.935	0.0088		FP	−0.982	0.0849
	GMM	−0.945	0.0056		GMM	−0.998	0.0071
	New	−0.939	0.0059		New	−0.998	0.0068

Simulation Study

$$n = 500,\ B = 2000$$

Table 2. Monte Carlo Means and Variances of the Estimators of θ under Model 2

$m(x)$	Method	Mean	Variance	$m(x)$	Method	Mean	Variance
Case 1	Full	−1.001	0.0025	Case 3	Full	−1.001	0.0025
	MAR	−1.041	0.0029		MAR	−1.052	0.0033
	FP	−0.999	0.0040		FP	−1.002	0.0058
	GMM	−1.001	0.0038		GMM	−1.001	0.0065
	New	−1.001	0.0038		New	−1.002	0.0059
Case 2	Full	−0.940	0.0029	Case 4	Full	−0.999	0.0021
	MAR	−0.992	0.0034		MAR	−1.045	0.0027
	FP	−0.929	0.0093		FP	−0.954	0.0741
	GMM	−0.940	0.0048		GMM	−0.999	0.0057
	New	−0.938	0.0048		New	−0.995	0.0055

Simulation Study

$$n = 500, \ B = 2000$$

Table 3. Monte Carlo Means and Variances of the Estimators of θ under Model 3

$m(x)$	Method	Mean	Variance	$m(x)$	Method	Mean	Variance
Case 1	Full	0.003	0.0023	Case 3	Full	0.000	0.0021
	MAR	−0.038	0.0027		MAR	−0.055	0.0025
	FP	0.005	0.0056		FP	0.000	0.0155
	GMM	0.009	0.0056		GMM	0.011	0.0156
	New	−0.004	0.0038		New	0.014	0.0049
Case 2	Full	0.059	0.0027	Case 4	Full	−0.000	0.0018
	MAR	0.005	0.0030		MAR	−0.048	0.0022
	FP	0.070	0.0100		FP	0.013	0.0306
	GMM	0.063	0.0059		GMM	0.012	0.0104
	New	0.060	0.0045		New	−0.006	0.0042

$$n = 200,\ B = 1000$$

	Model 1								
	m(x)	Method	Mean	Variance	m(x)	Method	Mean	Variance	
	Case 1	Full	-0.9991	0.00703	Case 3	Full	-1.0010	0.00658	
		MAR	-1.0192	0.01101		MAR	-1.0382	0.01213	
	code error	FP	-1.0331	0.00857	code error	FP	-1.0445	0.00876	
		GMM	-0.9976	0.01174		GMM	-0.9978	0.01962	
		New	-0.9990	0.01156		New	-0.9983	0.01784	
	Case 2	Full	-0.9401	0.00809	Case 4	Full	-1.0029	0.00620	
		MAR	-0.9379	0.02464		MAR	-1.0412	0.01100	
	code error	FP	-0.9851	0.00993	code error	FP	-1.0411	0.00825	
		GMM	-0.9405	0.01372		GMM	-0.9983	0.02114	
		New	-0.9343	0.01402		New	-0.9982	0.01945	

Simulation Study

$$n = 200, \; B = 1000$$

Model 2								
m(x)	Method	Mean	Variance	m(x)	Method	Mean	Variance	
Case 1	Full	-0.9995	0.00561	Case 3	Full	-0.9987	0.00565	
	MAR	-1.0201	0.00905		MAR	-1.0316	0.00970	
code error	FP	-1.0282	0.00681	code error	FP	-1.0337	0.00754	
	GMM	-1.0003	0.00923		GMM	-0.9998	0.02296	
	New	-1.0006	0.00924		New	-0.9992	0.01400	
Case 2	Full	-0.9421	0.00703	Case 4	Full	-1.0019	0.00518	
	MAR	-0.9225	0.11085		MAR	-1.0339	0.01566	
code error	FP	-0.9810	0.00868	code error	FP	-1.0339	0.00662	
	GMM	-0.9430	0.01173		GMM	-1.0005	0.01499	
	New	-0.9405	0.01183		New	-1.0002	0.01468	

$$n = 200,\ B = 1000$$

	Model 3								
	m(x)	Method	Mean	Variance	m(x)	Method	Mean	Variance	
	Case 1	Full	-0.0014	0.00562	Case 3	Full	0.0006	0.00606	
		MAR	0.0177	0.04137		MAR	-0.0113	0.02518	
	code error	FP	-0.0317	0.00670	code error	FP	-0.0401	0.00745	
		GMM	0.0139	0.02421		GMM	0.0258	0.04745	
		New	-0.0093	0.00880		New	-0.0150	0.01315	
	Case 2	Full	0.0577	0.00649	Case 4	Full	-0.0047	0.00453	
		MAR	0.1190	0.09568		MAR	-0.0191	0.04365	
	code error	FP	0.0156	0.00818	code error	FP	-0.0392	0.00555	
		GMM	0.0713	0.02640		GMM	0.0147	0.02982	
		New	0.0577	0.01058		New	-0.0148	0.01067	

n = 500, B = 2000

Table 5. Monte Carlo Variance, Mean of Variance Estimates, and Relative Bias for GMM Estimator and New Estimator under $m_1(x)$ Mean Structure With Model 1

	Parameter	GMM			New		
		Variance	$E(\hat{V})$	R.bias	Variance	$E(\hat{V})$	R.bias
Case 1	ϕ_0	0.0277	0.0277	−0.00	0.0259	0.0268	0.04
	ϕ_1	0.0229	0.0223	−0.03	0.0220	0.0227	0.04
	$\theta = E(Y)$	0.0049	0.0048	−0.01	0.0047	0.0049	0.03

Conclusion and Remarks

- New MLE method for nonignorable nonresponse.
- It is based on $f(y|x, \delta = 1)$ and the result is not sensitive. In the simulation we use normal distribution for $f(y|x, \delta = 1)$, but the resulting estimates are nearly unbiased.
- It is efficient since it is based on MLE approach. However, it doesn't necessarily satisfy the calibration constraints, so there is still room for improvement.
- It provides consistent estimates for the standard errors. Thus, we can test the null hypothesis that the response mechanism is ignorable. So, we can do some pretest procedure, and furthre investigation on this direction will be a topic of future research.

Question?