12-2012

Some theory for propensity-score-adjustment estimators in survey sampling

Jae Kwang Kim
Iowa State University, jkim@iastate.edu

Minsun Kim Riddles
Iowa State University

Follow this and additional works at: http://lib.dr.iastate.edu/stat_las_pubs

Part of the Design of Experiments and Sample Surveys Commons, and the Statistical Methodology Commons

The complete bibliographic information for this item can be found at http://lib.dr.iastate.edu/stat_las_pubs/107. For information on how to cite this item, please visit http://lib.dr.iastate.edu/howtocite.html.

This Article is brought to you for free and open access by the Statistics at Iowa State University Digital Repository. It has been accepted for inclusion in Statistics Publications by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
1. Introduction

Consider a finite population of size $N$, where $N$ is known. For each unit $i$, $y_i$ is the study variable and $x_i$ is the $q$-dimensional vector of auxiliary variables. The parameter of interest is the finite population mean of the study variable, $\theta = N^{-1} \sum_{i=1}^{N} y_i$. The finite population $F_N = \{(x_1', y_1), (x_2', y_2), \ldots, (x_N', y_N)\}$ is assumed to be a random sample of size $N$ from a superpopulation distribution $F(x, y)$. Suppose a sample of size $n$ is drawn from the finite population according to a probability sampling design. Let $w_i = \pi_i^{-1}$ be the design weight, where $\pi_i$ is the first-order inclusion probability of unit $i$ obtained from the probability sampling design. Under complete response, the finite population mean can be estimated by the Horvitz-Thompson (HT) estimator, $\hat{\theta}_{HT} = N^{-1} \sum_{i=1}^{n} w_i y_i$, where $A$ is the set of indices appearing in the sample.

In the presence of missing data, the HT estimator $\hat{\theta}_{HT}$ cannot be computed. Let $r$ be the response indicator variable that takes the value one if $y$ is observed and takes the value zero otherwise. Conceptually, as discussed by Fay (1992), Shao and Steel (1999), and Kim and Rao (2009), the response indicator can be extended to the entire population as $R_N = \{r_1, r_2, \ldots, r_N\}$, where $r_i$ is a realization of the random variable $r$. In this case, the complete-case (CC) estimator $\hat{\theta}_{CC} = \sum_{i \in A} w_i r_i y_i / \sum_{i \in A} w_i r_i$ converges in probability to $E(Y \mid r = 1)$. Unless the response mechanism is missing completely at random in the sense that $E(Y \mid r = 1) = E(Y)$, the CC estimator is biased. To correct for the bias of the CC estimator, if the response probability

$$p(x, y) = \Pr(r = 1 \mid x, y) \quad (1)$$

is known, then the weighted CC estimator $\hat{\theta}_{WCC} = N^{-1} \sum_{i \in A} w_i r_i y_i / p(x_i, y_i)$ can be used to estimate $\theta$. Note that $\hat{\theta}_{WCC}$ is unbiased because $E[E[\sum_{i \in A} w_i r_i y_i / p(x_i, y_i) \mid F_N] = E[\sum_{i=1}^{N} r_i y_i / p(x_i, y_i) \mid F_N] = \sum_{i=1}^{N} y_i$.

If the response probability (1) is unknown, one can postulate a parametric model for the response probability $p(x, y; \phi)$ indexed by $\phi \in \Omega$ such that $p(x, y) = p(x, y; \phi_0)$ for some $\phi_0 \in \Omega$. We assume that there exists a $\sqrt{n}$-consistent estimator $\phi$ of $\phi_0$ such that

$$\sqrt{n} (\phi - \phi_0) = O_p(1), \quad (2)$$

where $g_n = O_p(1)$ indicates $g_n$ is bounded in probability. Using $\phi$, we can obtain the estimated response probability by $\hat{p}_y = p(x, y; \phi)$, which is often called the propensity score (Rosenbaum and Rubin 1983). The propensity-score-adjusted (PSA) estimator can be constructed as

$$\hat{\theta}_{PSA} = \frac{1}{N} \sum_{i \in A} w_i \frac{r_i}{\hat{p}_i} y_i. \quad (3)$$


Despite the popularity of PSA estimators, asymptotic properties of PSA estimators have not received much attention in survey sampling literature. Kim and Kim (2007) used a Taylor expansion to obtain the asymptotic mean and variance of PSA estimators and discussed variance estimation. Da Silva and Opsomer (2006) and Da Silva and...
Opsomer (2009) considered nonparametric methods to obtain PSA estimators.

In this paper, we discuss optimal PSA estimators in the class of PSA estimators of the form (3) that use a $\sqrt{n}$-consistent estimator $\hat{\phi}$. Such estimators are asymptotically unbiased for $\theta$. Finding minimum variance PSA estimators among this particular class of PSA estimators is a topic of major interest in this paper.

Section 2 presents the main results. An optimal PSA estimator using an augmented propensity score model is proposed in Section 3. In Section 4, variance estimation of the proposed estimator is discussed. Results from two simulation studies can be found in Section 5 and concluding remarks are made in Section 6.

2. Main results

In this section, we discuss some asymptotic properties of PSA estimators. We assume that the response mechanism does not depend on $y$. Thus, we assume that

$$\Pr(r = 1 \mid x, y) = \Pr(r = 1 \mid x) = p(x; \phi_0)$$

for some unknown vector $\phi_0$. The first equality implies that the data are missing-at-random (MAR), as we always observe $x$ in the sample. Note that the MAR condition is assumed in the population model. In the second equality, we further assume that the response mechanism is known up to an unknown parameter $\phi_0$. The response mechanism is slightly different from that of Kim and Kim (2007), where we assume the following regularity conditions:

- **[C1]** The response mechanism satisfies (4), where $p(x; \phi)$ is continuous in $\phi$ and $\phi_0$. The responses are independent in the sense that $\text{Cov}(r_i, r_j \mid x) = 0$ for $i \neq j$. Also, $p(x; \phi) > c$ for all $i$ for some fixed constant $c > 0$.

- **[C2]** The solution to (6) exists and is unique almost everywhere. The function $h_i(\phi) = h(x_i; \phi)$ in (6) has a bounded fourth moment. Furthermore, the partial derivative $\partial(\hat{U}_h(\phi)) / \partial \phi$ is nonsingular for all $n$.

- **[C3]** The estimating function $\hat{U}_h(\phi)$ in (6) converges in probability to $U_h(\phi) = \sum_{i=1}^{N} (r_i - p_i(\phi)) h_i(\phi)$ uniformly in $\phi$. Furthermore, the partial derivative $\partial(\hat{U}_h(\phi)) / \partial \phi$ converges in probability to $\partial(\hat{U}_h(\phi)) / \partial \phi$ uniformly in $\phi$. The solution $\phi_0$ to $\hat{U}_h(\phi) = 0$ satisfies $N^{1/2}(\phi - \phi_0) = O_p(1)$ under the response mechanism.

Condition [C1] states the regularity conditions for the response mechanism. Condition [C2] is the regularity condition for the solution $\phi_0$ to (6). In Condition [C3], some regularity conditions are imposed on the estimating function $\hat{U}_h(\phi)$ itself. By [C2] and [C3], we can establish the $\sqrt{n}$-consistency (2) of $\hat{\phi}$.
Now, the following theorem deals with some asymptotic properties of the PSA estimator $\hat{\theta}_{\text{PSA},h}$.

**Theorem 1** If conditions [C1] - [C3] hold, then under the joint distribution of the sampling mechanism and the response mechanism, the PSA estimator $\hat{\theta}_{\text{PSA},h}$ satisfies

$$\sqrt{n}(\hat{\theta}_{\text{PSA},h} - \hat{\theta}_{\text{PSA},h}) = o_p(1),$$

where

$$\hat{\theta}_{\text{PSA},h} = \frac{1}{N} \sum_{i \in A} w_i \left[ p_i \ h_i \ \gamma_i^* + \frac{r_i}{p_i} (y_i - p_i \ h_i \ \gamma_i^*) \right].$$

(8)

which proves (7). The variance of $\hat{\theta}_{\text{PSA},h}$ can be derived as

$$V(\hat{\theta}_{\text{PSA},h}) = V(\hat{\theta}_{\text{HT}}) + \frac{1}{N^2} E \left[ \sum_{i \in A} w_i^2 \left( \frac{1}{p_i} - 1 \right) \right] (y_i - E(Y \mid x_i))^2$$

(14)

where the last equality follows because $y_i$ is conditionally independent of $E(Y \mid x_i) - p_i h_i \gamma_i^*$, conditioning on $x_i$. Since the last term in (14) is non-negative, the inequality in (9) is established. Furthermore, if $E(Y \mid x_i) = p_i h_i \alpha$ for some $\alpha$, then (10) holds and $E(\gamma_i^* \mid x_i) = \alpha$, by the definition of $\gamma_i^*$. Thus, $E(Y \mid x_i) - p_i h_i \gamma_i^* = -p_i h_i \alpha$, and $E(\gamma_i^* \mid x_i) = o_p(1)$, implying that the last term in (14) is negligible.

In (9), $V_i$ is the lower bound of the asymptotic variance of PSA estimators of the form (3) satisfying (6). Any PSA estimator that has the asymptotic variance $V_i$ in (9) is optimal in the sense that it achieves the lower bound of the asymptotic variance among the class of PSA estimators with $h$ satisfying (2). The asymptotic variance of optimal PSA estimators of $\theta$ is equal to $V_i$ in (9). The PSA estimator using the maximum likelihood estimator of $\phi_i$ does not necessarily achieve the lower bound of the asymptotic variance.

Condition (10) provides a way of constructing an optimal PSA estimator. First, we need an assumption for $E(Y \mid x)$, which is often called the outcome regression model. If the outcome regression model is a linear regression model of the form $E(Y \mid x) = \beta_0 + \beta_1 x$, an optimal PSA estimator of $\theta$ can be obtained by solving

$$\sum_{i \in A} w_i \frac{r_i}{p_i} (1, x_i) = \sum_{i \in A} w_i (1, x_i).$$

(15)
If we explicitly use a regression model for \( E(Y | x) \), it is possible to construct an estimator that has asymptotic variance (9) and is not necessarily a PSA estimator. For example, if we assume that

\[
E(Y | x) = m(x; \beta_0)
\]

(16)

for some function \( m(x; \cdot) \) known up to \( \beta_0 \), we can use the model (16) directly to construct an optimal estimator of the form

\[
\hat{\theta}_{\text{opt}} = \frac{1}{N \sum_{i = A} w_i} \left[ m(x_i; \hat{\beta}) + \frac{r_i}{p_i(\phi)} \{ y_i - m(x_i; \hat{\beta}) \} \right],
\]

(17)

where \( \hat{\beta} \) is a \( \sqrt{n} \)-consistent estimator of \( \beta_0 \) in the superpopulation model (16) and \( \phi \) is a \( \sqrt{n} \)-consistent estimator of \( \phi_0 \) computed by (6). The following theorem shows that the optimal estimator (17) achieves the lower bound in (9).

**Theorem 2** Let the conditions of Theorem 1 hold. Assume that \( \hat{\beta} \) satisfies \( \hat{\beta} = \beta_0 + O_p(n^{-1/2}) \). Assume that, in the superpopulation model (16), \( m(x; \beta) \) has continuous first-order partial derivatives in an open set containing \( \beta_0 \). Under the joint distribution of the sampling mechanism, the response mechanism, and the superpopulation model (16), the estimator \( \hat{\theta}_{\text{opt}} \) in (17) satisfies

\[
\sqrt{n} (\hat{\theta}_{\text{opt}} - \theta^*_o) = o_p(1),
\]

where

\[
\theta^*_o = \frac{1}{N \sum_{i = A} w_i} \left[ m(x_i; \beta_0) + \frac{r_i}{p_i} \{ y_i - m(x_i; \beta_0) \} \right],
\]

(18)

\[
p_i = p_i(\phi_0), \quad \text{and} \quad V(\hat{\theta}_{\text{opt}}) = \text{equal to } V_r \text{ in (9)}.
\]

**Proof.** Define \( \hat{\theta}^*_{\text{opt}}(\beta, \phi) = N^{-1} \sum_{i = A} w_i \{ m(x_i; \beta) + r_i p_i^{-1} \{ y_i - m(x_i; \beta) \} \} \). Note that \( \hat{\theta}_{\text{opt}} \) in (17) can be written as \( \hat{\theta}_{\text{opt}} = \hat{\theta}^*_{\text{opt}}(\beta, \phi) \). Since

\[
\frac{\partial}{\partial \beta} \hat{\theta}^*_{\text{opt}}(\beta, \phi) = \frac{1}{N \sum_{i = A} w_i} \left[ \hat{m}(x_i; \phi) - \frac{r_i}{p_i} \hat{m}(x_i; \beta) \right],
\]

where \( \hat{m}(x; \beta) = \hat{m}(x_i; \beta) / \hat{\phi} \), and

\[
\frac{\partial}{\partial \phi} \hat{\theta}^*_{\text{opt}}(\beta, \phi) = \frac{1}{N \sum_{i = A} w_i} \left[ \hat{r}_i z_i(\phi) \{ y_i - m(x_i; \beta) \} \right],
\]

where \( z_i(\phi) = \partial \{ p_i^{-1}(\phi) \} / \partial \phi \). We have \( E[\partial \{ \hat{\theta}^*_{\text{opt}}(\beta, \phi) \} / \partial \phi] | \beta = \beta_0, \phi = \phi_0 \) = 0 and the condition of Randles (1982) is satisfied. Thus, \( \hat{\theta}^*_{\text{opt}}(\beta, \phi) = \hat{\theta}_{\text{opt}}(\beta_0, \phi_0) + o_p(n^{-1/2}) = \theta^*o + o_p(n^{-1/2}) \) and the variance of \( \hat{\theta}^*_{\text{opt}} \) is equal to \( V_r \), the lower bound of the asymptotic variance.

The (asymptotic) optimality of the estimator in (17) is justified under the joint distribution of the response model (4) and the superpopulation model (16). When both models are correct, \( \hat{\theta}_{\text{opt}} \) is optimal and the choice of \( (\hat{\beta}, \hat{\phi}) \) does not affect the efficiency of the \( \hat{\theta}_{\text{opt}} \) as long as \( (\hat{\beta}, \hat{\phi}) \) is \( \sqrt{n} \)-consistent. Robins, Rotnitzky and Zhao (1994) also advocated using \( \hat{\theta}_{\text{opt}} \) in (17) under simple random sampling.

**Remak 1** When the response model is correct and the superpopulation model (16) is not necessarily correct, the choice of \( \hat{\beta} \) does affect the efficiency of the optimal estimator. Cao, Tsatis and Davidian (2009) considered optimal estimation when only the response model is correct. Using Taylor linearization, the optimal estimator in (17) with \( \hat{\phi} \) satisfying (6) is asymptotically equivalent to

\[
\hat{\theta}(\beta) = \sum_{i \in A} w_i \left[ m(x_i; \beta) + \frac{r_i}{p_i} \{ y_i - m(x_i; \beta) \} - \left( \frac{r_i}{p_i} - 1 \right) c^*_p p_i h_i \right],
\]

where \( c^*_p \) is the probability limit of \( \hat{c}_p = \{ \sum_{i \in A} w_i r_i z_i(\hat{\phi}) \} \hat{p}_i h_i(\hat{\phi})^{-1} \sum_{i \in A} w_i r_i z_i(\hat{\phi}) \{ y_i - m(x_i; \beta) \} \) and \( z_i(\phi) = \partial \{ p_i^{-1}(\phi) \} / \partial \phi \). The asymptotic variance is then equal to

\[
V\{\hat{\theta}(\beta)\} = \sum_{i \in A} w_i^2 \frac{1 - \hat{p}_i}{\hat{p}_i^2} \{ y_i - m(x_i; \beta) - c^*_p p_i h_i(\hat{\phi}) \}^2.
\]

Thus, an optimal estimator of \( \beta \) can be computed by finding \( \hat{\beta} \) that minimizes

\[
Q(\beta) = \sum_{i \in A} w_i^2 \frac{1 - \hat{p}_i}{\hat{p}_i^2} \{ y_i - m(x_i; \beta) - c^*_p p_i h_i(\hat{\phi}) \}^2.
\]

The resulting estimator is design-optimal in the sense that it minimizes the asymptotic variance under the response model.

### 3. Augmented propensity score model

In this section, we consider optimal PSA estimation. Note that the optimal estimator \( \hat{\theta}_{\text{opt}} \) in (17) is not necessarily written as a PSA estimator form in (3). It is in the PSA estimator form if it satisfies \( \sum_{i \in A} w_i r_i \hat{p}_i^{-1} m(x_i; \beta) = \sum_{i \in A} w_i m(x_i; \beta) \). Thus, we can construct an optimal PSA estimator by including \( m(x_i; \beta) \) in the model for the propensity score. Specifically, given \( \hat{m}_i = m(x_i; \beta) \), \( \hat{p}_i = p_i(\hat{\phi}) \) and \( \hat{h}_i = h_i(\hat{\phi}) \), where \( \hat{\phi} \) is obtained from (6), we augment the response model by

\[
p_i^*(\hat{\phi}, \lambda) = \frac{\hat{p}_i}{\hat{p}_i + (1 - \hat{p}_i) \exp(\lambda_0 + \lambda_1 \hat{m}_i)},
\]

(18)
where \( \lambda = (\lambda_0, \lambda_1)' \) is the Lagrange multiplier which is used to incorporate the constraint. If \( (\lambda_0, \lambda_1)' = 0 \), then \( p_i(\hat{\phi}, \hat{\lambda}) = \hat{p}_i \). The augmented response probability \( p_i(\hat{\phi}, \lambda) \) always takes values between 0 and 1. The augmented response probability model (18) can be derived by minimizing the Kullback-Leibler distance \( \sum_{i \in A} w_i q_i' \log(q_i'/q_i) \), where \( q_i' = (1 - p_i')/\hat{p}_i' \) and \( q_i = (1 - \hat{p}_i)/\hat{p}_i \), subject to the constraint \( \sum_{i \in A} w_i (r_i/p_i)'(1, \hat{m}_i) = \sum_{i \in A} w_i (1, \hat{m}_i) \).

Using (18), the optimal PSA estimator is computed by
\[
\hat{\theta}_{\text{PSA}}^* = \frac{1}{N} \sum_{i \in A} w_i \frac{r_i}{p_i(\hat{\phi}, \lambda)} y_i,
\]
where \( \hat{\lambda} \) satisfies
\[
\sum_{i \in A} w_i \frac{r_i}{p_i(\hat{\phi}, \lambda)} (1, \hat{m}_i) = \sum_{i \in A} w_i (1, \hat{m}_i).
\]

Under the response model (4), it can be shown that
\[
\hat{\theta}_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i \left\{ \hat{b}_0 + \hat{b}_1 \hat{m}_i + \frac{r_i}{\hat{p}_i} (y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i) \right\} + o_p(n^{-1/2}),
\]
where
\[
\left( \begin{array}{c} \hat{b}_0 \\ \hat{b}_1 \end{array} \right) = \left( \sum_{i \in A} w_i r_i \left( 1 + 1 \right)^{-1} \right)^{-1} \left( \sum_{i \in A} w_i r_i \left( 1 \right)^{-1} \right) \left( y_i - \hat{m}_i \right)
\]
and
\[
\sum_{i \in A} w_i r_i \left( 1 \right)^{-1} \left( y_i - \hat{m}_i \right).
\]

Furthermore, by the argument for Theorem 1, we can establish that
\[
\hat{\theta}_{\text{PSA}}^* = \frac{1}{N} \sum_{i \in A} w_i \left\{ b_0 + b_1 \hat{m}_i + \gamma_{b2} p_i h_i \right.
\]
\[
+ \frac{r_i}{p_i} \left( y_i - b_0 - b_1 \hat{m}_i - \gamma_{b2} p_i h_i \right)
\]
\[
+ o_p(n^{-1/2}),
\]
where \( (b_0, b_1, \gamma_{b2}) \) is the probability limit of \( (\hat{b}_0, \hat{b}_1, \gamma_{b2}') \) with
\[
\left( \begin{array}{c} \gamma_{b2} \\ \gamma_{b2}' \end{array} \right) = \left( \sum_{i \in A} w_i r_i z_i(\hat{\phi}) \right)^{-1} \left( \sum_{i \in A} w_i r_i z_i(\hat{\phi}) (y_i - \hat{b}_0 - \hat{b}_1 \hat{m}_i) \right)
\]
and the effect of estimating \( \phi_0 \) in \( \hat{p}_i = p(x_i; \hat{\phi}) \) can be safely ignored.

Note that, under the response model (4), \( (\hat{\phi}, \hat{\lambda}) \) in (19) converges in probability to \( (\phi_0, 0) \), where \( \phi_0 \) is the true parameter in (4). Thus, the propensity score from the augmented model converges to the true response probability. Because \( \hat{\lambda} \) converges to zero in probability, the choice of \( \hat{\beta} \) in \( \hat{m}_i = m(x_i; \hat{\beta}) \) does not play a role for the asymptotic unbiasedness of the PSA estimator. The asymptotic variances are changed for different choices of \( \hat{\beta} \).

Under the superpopulation model (16), \( \hat{b}_0 + \hat{b}_1 \hat{m}_i \rightarrow E(Y | x_i) \) in probability. Thus, the optimal PSA estimator in (19) is asymptotically equivalent to the optimal estimator in (17). Incorporating \( \hat{m}_i \) into the calibration equation to achieve optimality is close in spirit to the model-calibration method proposed by Wu and Sitter (2001).

4. Variance estimation

We now discuss variance estimation of PSA estimators under the assumed response model. Singh and Folsom (2000) and Kott (2006) discussed variance estimation for certain types of PSA estimators. Kim and Kim (2007) discussed variance estimation when the PSA estimator is computed with the maximum likelihood method. We consider variance estimation for the PSA estimator of the form (3) where \( \hat{p}_i = p_i(\hat{\phi}) \) is constructed to satisfy (6) for some \( h_i(\hat{\phi}) = h(x_i; \hat{\phi}, \hat{\beta}) \), where \( \hat{\beta} \) is obtained using the postulated superpopulation model. Let \( \beta^* \) be the probability limit of \( \hat{\beta} \) under the response model. Note that \( \beta^* \) is not necessarily equal to \( \beta_0 \) in (16) since we are not assuming that the postulated superpopulation model is correctly specified in this section.

Using the argument for the Taylor linearization (13) used in the proof of Theorem 1, the PSA estimator satisfies
\[
\hat{\theta}_{\text{PSA}} = \frac{1}{N} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*) + o_p(n^{-1/2}),
\]
where
\[
\eta_i(\phi_0, \beta^*) = p_i(\hat{\phi}) h_i'(\phi, \beta^*) \gamma_{b}^* + \frac{r_i}{p_i(\hat{\phi})} \left( y_i - p_i(\hat{\phi}) h_i(\phi, \beta^*) \gamma_{b}^* \right),
\]
and
\[
\eta_i(\phi_0, \beta^*) = h(x_i; \phi, \beta) \text{ and } \gamma_{b}^* \text{ is defined as in (8) with } h_i \text{ replaced by } h_i(\phi_0, \beta^*). \text{ Since } p_i(\hat{\phi}) \text{ satisfies (6) with } h_i(\hat{\phi}) = h(x_i; \hat{\phi}, \hat{\beta}), \hat{\theta}_{\text{PSA}} = N^{-1} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*) \text{ holds and the linearization in (23) can be expressed as } N^{-1} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*) = N^{-1} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*) + o_p(n^{-1/2}). \text{ Thus, if } (x_i, y_i, r_i) \text{ are independent and identically distributed (IID), then } \eta_i(\phi_0, \beta^*) \text{ are IID even though } \eta_i(\phi_0, \beta^*) \text{ are not necessarily IID. Because } \eta_i(\phi_0, \beta^*) \text{ are IID, we can apply the standard complete sample method to estimate the variance of } \hat{\eta}_{\text{PSA}} = N^{-1} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*), \text{ which is asymptotically equivalent to the variance of } \hat{\theta}_{\text{PSA}} = N^{-1} \sum_{i \in A} w_i \eta_i(\phi_0, \beta^*). \text{ See Kim and Rao (2009).}

To derive the variance estimator, we assume that the variance estimator \( \hat{\sigma} = N^{-2} \sum_{i \in A} \sum_{j \in A} \Omega_{ij} g_i g_j \) satisfies
\[
\hat{V}/V(\hat{g}_{HT} | \mathcal{F}_N) = 1 + o_n(1)
\] for some \( \Omega_{ij} \) related to the joint inclusion probability, where \( \hat{g}_{HT} = N^{-1}\sum_{i=1}^{N} w_i g_{ij} \), for any \( g \) with a finite second moment and \( V(g_{HT} | \mathcal{F}_N) = N^{-2}\sum_{i=1}^{N} \sum_{j=1}^{N} \Omega_{ij} g_{ij} g_{ij} \), for some \( \Omega_{ij} \). We also assume that
\[
\sum_{j=1}^{N} |\Omega_{ij}| = O(n^{-1}N).
\] (25)

To obtain the total variance, the reverse framework of Fay (1992), Shao and Steel (1999), and Kim and Rao (2009) is considered. In this framework, the finite population is first divided into two groups, a population of respondents and a population of nonrespondents. Given the population, the sample is selected according to a probability sampling design. Thus, selection of the population respondents is treated as the first-phase sampling and the selection of the sample respondents from the whole finite population is treated as the second-phase sampling in the reverse framework. The total variance of the population respondents is treated as the second-phase variance term. The conditional variance term \( HT_1(\cdot, \cdot) \) is also consistent for \( H_1^* \) and \( V_1^* \) is defined after (27). Therefore, the variance estimation of the optimal PSA estimator can be constructed as in (29).

\[
\text{Remark 2 The variance estimation of the optimal PSA estimator with augmented propensity model (18) with } (\hat{\phi}, \hat{\lambda}) \text{ satisfying (20) can be derived by (29) using } \hat{N}_1 = \hat{b}_0 + \hat{b}_1 \hat{m} + \hat{\gamma}_{12} \hat{h} + r_i \hat{p}_i \hat{y}_i - \hat{b}_0 \hat{m} - \hat{\gamma}_{12} \hat{h} \text{ where } (\hat{b}_0, \hat{b}_1) \text{ and } \hat{\gamma}_{12} \text{ are defined in (21) and (22), respectively.}
\]

\section{Simulation study}

\subsection{Study one}

Two simulation studies were performed to investigate the properties of the proposed method. In the first simulation, we generated a finite population of size \( N = 10,000 \) from the following multivariate normal distribution:
\[
\begin{pmatrix}
x_1 \\
x_2 \\
e
\end{pmatrix} \sim N \left( \begin{pmatrix}
2 \\
-1 \\
0.5
\end{pmatrix}, \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} \right).
\]

The variable of interest \( y \) was constructed as \( y = 1 + x_1 + e \). We also generated response indicator variables \( r_i \) independently from a Bernoulli distribution with probability
\[
p_i = \frac{\exp(2 + x_{2i})}{1 + \exp(2 + x_{2i})}.
\] (30)

From the finite population, we used simple random sampling to select two samples of size, \( n = 100 \) and \( n = 400 \), respectively. We used \( B = 5,000 \) Monte Carlo samples in the simulation. The average response rate was about 69.6%.

To compute the propensity score, a response model of the form
\[
p(x; \phi) = \frac{\exp(\phi_0 + \phi_1 x)}{1 + \exp(\phi_0 + \phi_1 x)}
\] (31)

was postulated and an outcome regression model of the form
\[
m(x; \beta) = \beta_0 + \beta_1 x_1
\] was postulated to obtain the optimal PSA estimators. Thus, both models are correctly specified. From each sample, we computed four estimators of \( \theta = N^{-1}\sum_{i=1}^{N} y_i \):
1. (PSA-MLE): PSA estimator in (3) with \( \hat{p}_i = p_i(\hat{\phi}) \) and \( \hat{\phi} \) being the maximum likelihood estimator of \( \phi \).

2. (PSA-CAL): PSA estimator in (3) with \( \hat{p}_i \) satisfying the calibration constraint (15) on \( (1, x_{2i}) \).

3. (AUG): Augmented PSA estimator in (19).

4. (OPT): Optimal estimator in (17).

In the augmented PSA estimators, \( \hat{\phi} \) was computed by the maximum likelihood method. Under model (30), the maximum likelihood estimator of \( \phi = (\phi_0, \phi_1)' \) was computed by solving (6) with \( h_i(\phi) = (1, x_{2i})' \). Parameter \( (\beta_0, \beta_1) \) for the outcome regression model was computed using ordinary least squares, regressing \( y \) on \( x \). In addition to the point estimators, we also computed the variance estimators of the point estimators. The variance estimators of the PSA estimators were computed using the pseudo-values in (24) and the \( h_i(\phi) \) corresponding to each estimator. For the augmented PSA estimators, the pseudo-values were computed by the method in Remark 2.

Table 1 presents the Monte Carlo biases, variances, and mean square errors of the four point estimators and percent relative biases (R.B.) and \( \tau \)-statistics of the variance estimators based on 5,000 Monte Carlo samples in Table 1 is the test statistic for testing the zero bias of the variance estimator. See Kim (2004).

Based on the simulation results in Table 1, we have the following conclusions.

1. All of the PSA estimators are asymptotically unbiased because the response model (30) is correctly specified. The PSA estimator using the calibration method is slightly more efficient than the PSA estimator using the maximum likelihood estimator, because the last term of (14) is smaller for the calibration method as the predictor for \( E(Y | x_i) = \beta_0 + \beta_1 x_{2i} \) is better approximated by a linear function of \( (1, x_{2i}) \) than by a linear function of \( (\hat{p}_1, \hat{p}_1, x_{2i}) \).

2. The augmented PSA estimator is more efficient than the direct PSA estimator (3). The augmented PSA estimator is constructed by using the correctly specified regression model (31) and so it is asymptotically equivalent to the optimal PSA estimator in (17).

3. Variance estimators are all approximately unbiased. There are some modest biases in the variance estimators of the PSA estimators when the sample size is small (\( n = 100 \)).

5.2 Study two

In the second simulation study, we further investigated the PSA estimators with a non-linear outcome regression model under an unequal probability sampling design. We generated two stratified finite populations of \( (x, y) \) with four strata \( (h = 1, 2, 3, 4) \), where \( x_{hi} \) were independently generated from a normal distribution \( N(1, 1) \) and \( y_{hi} \) were dichotomous variables that take values of 1 or 0 from a Bernoulli distribution with probability \( p_{1yhi} \) or \( p_{2yhi} \). Two different probabilities were used for two populations, respectively:

1. Population 1 (Pop1):
   \[ p_{1yhi} = 1 / \left[ 1 + \exp(0.5 - 2x) \right] \]

2. Population 2 (Pop2):
   \[ p_{2yhi} = 1 / \left[ 1 + \exp(0.25(x - 1.5)^2 - 1.5) \right] \]

In addition to \( x_{hi} \) and \( y_{hi} \), the response indicator variables \( r_{hi} \) were generated from a Bernoulli distribution with probability \( p_{hi} = 1 / \left[ 1 + \exp(-1.5 + 0.7x_{hi}) \right] \). The sizes of the four strata were \( N_1 = 1,000 \), \( N_2 = 2,000 \), \( N_3 = 3,000 \), and \( N_4 = 4,000 \), respectively. In each of the two sets of finite population, a stratified sample of size \( n = 400 \) was independently generated without replacement, where a simple random sample of size \( n_h = 100 \) was selected from each stratum. We used \( B = 5,000 \) Monte Carlo samples in this simulation. The average response rate was about 67%.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Monte Carlo bias, variance and mean square error(MSE) of the four point estimators and percent relative biases (R.B.) and ( \tau )-statistics(( \tau )-stat) of the variance estimators based on 5,000 Monte Carlo samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Method</td>
<td>Bias</td>
</tr>
<tr>
<td>PSA-MLE</td>
<td>-0.01</td>
</tr>
<tr>
<td>PSA-CAL</td>
<td>-0.01</td>
</tr>
<tr>
<td>AUG</td>
<td>0.00</td>
</tr>
<tr>
<td>OPT</td>
<td>0.00</td>
</tr>
<tr>
<td>--------</td>
<td>---</td>
</tr>
<tr>
<td>PSA-MLE</td>
<td>-0.01</td>
</tr>
<tr>
<td>PSA-CAL</td>
<td>-0.01</td>
</tr>
<tr>
<td>AUG</td>
<td>0.00</td>
</tr>
<tr>
<td>OPT</td>
<td>0.00</td>
</tr>
</tbody>
</table>
To compute the propensity score, a response model of the form
\[ p(x; \phi) = \frac{\exp(\phi_0 + \phi_1 x)}{1 + \exp(\phi_0 + \phi_1 x)} \]
was postulated for parameter estimation. To obtain the augmented PSA estimator, a model for the variable of interest of the form
\[ m(x; \beta) = \frac{\exp(\beta_0 + \beta_1 x)}{1 + \exp(\beta_0 + \beta_1 x)} \]  
(32)
was postulated. Thus, model (32) is a true model under (Pop1), but it is not a true model under (Pop2).

We computed four estimators:
1. (PSA-MLE): PSA estimator in (3) using the maximum likelihood estimator of $\hat{\phi}$.
2. (PSA-CAL): PSA estimator in (3) with $\hat{p}_i$ satisfying the calibration constraint (15) on $(1, x)$.
3. (AUG-1): Augmented PSA estimator $\hat{\theta}_{PSA}$ in (19) with $\hat{\beta}$ computed by the maximum likelihood method.
4. (AUG-2): Augmented PSA estimator $\hat{\theta}_{PSA}$ in (19) with $\hat{\beta}$ computed by the method of Cao et al. (2009) discussed in Remark 1.

We considered the the augmented PSA estimator in (19) with $\hat{p}_i = p_i(\hat{\phi})$, where $\hat{\phi}$ is the maximum likelihood estimator of $\phi$. The first augmented PSA estimator (AUG-1) used $\hat{m}_i = m(x_i; \hat{\beta})$ with $\hat{\beta}$ found by solving $\sum_{i \in h} \sum_{u \in h} w_{hi} r_{hi} \{y_{hi} - m(x_{hi}; \hat{\beta})\}(1, x_{hi}) = 0$, where $A_h$ is the set of indices appearing in the sample for stratum $h$ and $w_{hi}$ is the sampling weight of unit $i$ for stratum $h$.

Table 2 presents the simulation results for each method. In each population, the augmented PSA estimator shows some improvement comparing to the PSA estimator using the maximum likelihood estimator of $\phi$ or the calibration estimator of $\phi$ in terms of variance. Under (Pop1), since model (32) is true, there is essentially no difference between the augmented PSA estimators using different methods of estimating $\beta$. However, under (Pop2), where the assumed outcome regression model (32) is incorrect, the augmented PSA estimator with $\hat{\beta}$ computed by the method of Cao et al. (2009) results in slightly better efficiency, which is consistent with the theory in Remark 1. Variance estimates are approximately unbiased in all cases in the simulation study.

6. Conclusion

We have considered the problem of estimating the finite population mean of $y$ under nonresponse using the propensity score method. The propensity score is computed from a parametric model for the response probability, and some asymptotic properties of PSA estimators are discussed. In particular, the optimal PSA estimator is derived with an additional assumption for the distribution of $y$. The propensity score for the optimal PSA estimator can be implemented by the augmented propensity model presented in Section 3. The resulting estimator is still consistent even when the assumed outcome regression model fails to hold.

We have restricted our attention to missing-at-random mechanisms in which the response probability depends only on the always-observed $x$. If the response mechanism also depends on $y$, PSA estimation becomes more challenging. PSA estimation when missingness is not at random is beyond the scope of this article and will be a topic of future research.

Acknowledgements

The research was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. The authors wish to thank F. Jay Breidt, three anonymous referees, and the associate editor for their helpful comments.

Table 2
Monte Carlo bias, variance and mean square error of the four point estimators and percent relative biases (R.B.) and $t$-statistics of the variance estimators, based on 5,000 Monte Carlo samples

<table>
<thead>
<tr>
<th>Population</th>
<th>Method</th>
<th>Bias</th>
<th>Variance</th>
<th>MSE</th>
<th>R.B. (%)</th>
<th>$t$-stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop1</td>
<td>(PSA-MLE)</td>
<td>0.00</td>
<td>0.000750</td>
<td>0.000762</td>
<td>-1.13</td>
<td>-0.57</td>
</tr>
<tr>
<td></td>
<td>(PSA-CAL)</td>
<td>0.00</td>
<td>0.000762</td>
<td>0.000769</td>
<td>-1.45</td>
<td>-0.72</td>
</tr>
<tr>
<td></td>
<td>(AUG-1)</td>
<td>0.00</td>
<td>0.000745</td>
<td>0.000757</td>
<td>-1.73</td>
<td>-0.86</td>
</tr>
<tr>
<td></td>
<td>(AUG-2)</td>
<td>0.00</td>
<td>0.000745</td>
<td>0.000757</td>
<td>-1.83</td>
<td>-0.91</td>
</tr>
<tr>
<td>Pop2</td>
<td>(PSA-MLE)</td>
<td>0.00</td>
<td>0.000824</td>
<td>0.000826</td>
<td>0.29</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>(PSA-CAL)</td>
<td>0.00</td>
<td>0.000829</td>
<td>0.000835</td>
<td>-0.94</td>
<td>-0.46</td>
</tr>
<tr>
<td></td>
<td>(AUG-1)</td>
<td>0.00</td>
<td>0.000822</td>
<td>0.000823</td>
<td>-0.71</td>
<td>-0.35</td>
</tr>
<tr>
<td></td>
<td>(AUG-2)</td>
<td>0.00</td>
<td>0.000820</td>
<td>0.000821</td>
<td>-0.61</td>
<td>-0.30</td>
</tr>
</tbody>
</table>


