The Journal of Applied Science **Vol. 18 No. 2: 1-10 [2019]**
วารสารวิทยาศาสตร์ประยุกต์ doi: 10.14416/j.appsci.2019.07.001
ISSN 1513-7805 Printed in Thailand ISSN 2586-9663 (Online)

Research Article

A new variance estimator for unequal probability sampling without replacement in the presence of non-response

Nuanpan Lawson[1]* and Chugiat Ponkaew[1]

[1]Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok 10800, Thailand
***E-mail**: nuanpan.n@sci.kmutnb.ac.th
Received: 13/05/2019; Revised: 17/06/2019; Accepted: 22/07/2019

Abstract

A new variance estimator for estimating the population total has been proposed under unequal probability sampling without replacement in the presence of nonresponse. The new variance estimator does not require the joint inclusion probability in the estimation under the reverse framework where the sampling fraction is negligible and equal response probabilities for all units. The efficiency of the new estimator is compared with the existing estimators via a simulation study.

Keywords: variance estimator, response probabilities, reverse framework, joint inclusion probability

Introduction

Lack of complete information in sample surveys effects planning, decision making in business and industry. Ignoring missing data can lead to bias in estimation and may lead to incorrect conclusions by drawing the results based on available incomplete data. In the case of full response, Horvitz & Thompson (1952) proposed a well-known population total estimator for unequal probability sampling without replacement. The Horvitz & Thompson's estimator (1952) is given by,

$$\hat{Y}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i}, \tag{1}$$

where π_i is the first order inclusion probability of unit i in sample s, $\pi_i = P(i \in s)$.

The variance estimator of $\hat{Y}_{HT}$, $\hat{V}(\hat{Y}_{HT})$ is given by,

$$\hat{V}(\hat{Y}_{HT}) = \sum_{i \in s} \left(\frac{1-\pi_i}{\pi_i^2} \right) y_i^2 + \sum_{i \in s} \sum_{j \setminus \{i\} \in s} \left(\frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j} \right) y_i y_j, \tag{2}$$

where π_{ij} is the joint inclusion probability of units i and j in s of sample size n, $\pi_{ij} = P(i, j \in s)$. The joint inclusion probability is not always known and therefore later Hájek (1981) suggested an asymptotic approach to estimate the joint inclusion probability of the

 doi: 10.14416/j.appsci.2019.07.001

variance of the population total estimator. However, Berger (1998) stated that the Hájek approach can be used to estimate the value of π_{ij} only for high entropy sampling and it requires π_i for all $i \in U$.

When nonresponse occurs, Särndal & Lundström (2005) proposed a new population total estimator created by adjusting the Horvitz & Thompson estimator using the inverse of the response probability under a two-phrase framework. The estimated value of the variance of the new estimator was also proposed. However, the joint inclusion probabilities are usually unknown in practice. The Särndal & Lundström estimator is given by,

$$\hat{Y}_{SL} = \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i} , \tag{3}$$

where r_i is the response indicator variable of y_i and p_i is the response probability.

The variance of $\hat{Y}_{SL}$, $V\left(\hat{Y}_{SL}\right)$ is given by,

$$V\left(\hat{Y}_{SL}\right) = \frac{1}{2} \sum_{i \in U} \sum_{j \setminus i \in U} (\pi_i \pi_j - \pi_{ij}) (\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j})^2 + \sum_i \frac{y_i^2}{\pi_i} (\frac{1}{p_i} - 1) . \tag{4}$$

Lawson (2017) proposed a new estimator for estimating the population total along with a variance estimator for estimating the population total under unequal probability sampling with replacement when nonresponse occurs. A Taylor series is considered for estimating the variance of the population total estimator under the reverse framework when the response probabilities are uniform, $p_i = p$ for all units i in the population U and negligible sampling fraction. The response probabilities are not required to estimate variance in the Lawson's estimator. The Lawson (2017) estimator for estimating the population total is given by,

$$\hat{Y}_{NR} = N\hat{\bar{Y}}_{NR} = N \sum_{i \in s} \frac{r_i y_i}{\pi_i} \Big/ \sum_{i \in s} \frac{r_i}{\pi_i} , \tag{5}$$

where $\hat{\bar{Y}}_{NR} = \sum_{i \in s} \frac{r_i y_i}{\pi_i} \Big/ \sum_{i \in s} \frac{r_i}{\pi_i}$.

The variance of $\hat{Y}_{NR}$ is given by,

$$\begin{aligned} V(\hat{Y}_{NR}) &\approx n \sum_{i \in U} E_R(o_i^2) P_i - \frac{1}{n} \left(\sum_{i \in U} E_R(o_i) P_i \right)^2 \\ &= \frac{n}{p} \sum_{i \in U} \frac{1}{\pi_i^2} (y_i - \bar{Y})^2 P_i - \frac{1}{n} \left(\sum_{i \in U} \frac{1}{\pi_i} (y_i - \bar{Y}) P_i \right)^2 , \end{aligned} \tag{6}$$

 doi: 10.14416/j.appsci.2019.07.001

where $o_i = \frac{N}{\sum_{i\in U} r_i}\frac{r_i}{\pi_i}(y_i - \bar{Y}_{NR})$, $\bar{Y}_{NR} = \frac{\sum_{i\in U} r_i y_i}{\sum_{i\in U} r_i}$, $E_R(\cdot)$ is the expectation operator respect to nonresponse mechanism and P_i is the probability of the selection of unit i th in the population at any given time and it is defined by, $P_i = \frac{k_i}{K}$, where $K = \sum_{i\in U} k_i$.

The estimated variance of $\hat{Y}_{NR}$ is defined by,

$$\begin{aligned}\hat{V}(\hat{Y}_{NR}) &\approx \frac{n}{n-1}\left[\sum_{i\in s}\hat{o}_i^2 - \frac{1}{n}\left(\sum_{i\in s}\hat{o}_i\right)^2\right], \\ &= \frac{N^2}{\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)^2}\frac{n}{n-1}\sum_{i\in s}\frac{r_i}{\pi_i^2}\left(y_i - \hat{\bar{Y}}_{NR}\right)^2, \end{aligned} \tag{7}$$

where $\hat{o}_i = \frac{N}{\sum_{i\in s}\frac{r_i}{\pi_i}}\frac{r_i}{\pi_i}(y_i - \hat{\bar{Y}}_{NR})$.

Recently Ponkaew & Lawson (2019) proposed a new variance estimator for estimating population total following Särndal & Lundström (2005) in the presence of nonresponse under unequal probability sampling without replacement under a reverse framework. The Ponkaew & Lawson (2019) estimator is a linear estimator defined by,

$$\hat{Y}_r^{(1)} = \sum_{i\in s}\frac{r_i y_i}{\pi_i p}. \tag{8}$$

The estimated variance of the Ponkaew & Lawson (2019) estimator is considered in two situations where p is known and where p is unknown. The estimated value of the Ponkaew & Lawson (2019) variance estimators are defined by,

$$\hat{V}(\hat{Y}_r^{(1)}) = \frac{1}{p^2}\left[\sum_{i\in s}\frac{(1-\pi_i)}{\pi_i^2}r_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\breve{D}_{ij} r_i r_j y_i y_j\right], \tag{9}$$

$$\hat{V}(\hat{Y}_r^{(1)}) = \left(\frac{\sum_{i\in s}\frac{1}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2\left[\sum_{i\in s}\frac{(1-\pi_i)}{\pi_i^2}r_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\breve{D}_{ij} r_i r_j y_i y_j\right], \tag{10}$$

where $\breve{D}_{ij} = \frac{D_{ij}}{\pi_{ij}} = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$.

However, the Ponkaew & Lawson (2019) estimator requires that the joint inclusion probabilities π_{ij} which are usually unknown in practice, be known (Berger, 2003).

In this paper, a new variance of the population total estimator under unequal probability sampling without replacement is proposed following Lawson (2017) by using an idea of Ponkaew & Lawson (2019). The proposed variance estimator does not require that the joint inclusion probabilities under the reverse framework when the response probabilities are uniform and sampling fraction is negligible to be known. A simulation study will be used to compare the efficiency of the new estimator against the existing estimators.

Methods

Suppose that a sample s of size n is selected with a probability proportional to size sampling without replacement (PPSWOR) from a finite population $U=\{1,2,\ldots,N\}$ of size N. Let $Y=\sum_{i\in U} y_i$ be the population total of study variable y. Let r_i denote the response indicator variable of y_i where r_i is equal to 1 if y_i is observed; otherwise $r_i=0$ and let p_i denote the response probability, $p_i=P(r_i=1)$. Let $\hat{Y}=\sum_{i\in s} o_i$ denote the population total estimator where o_i is the function of y_i, $o_i=f(y_i)$.

Under the reverse framework with unequal probability sampling with replacement when the response probabilities are uniform and the sampling fraction is negligible, Lawson (2017) showed that the variance of $\hat{Y}=\sum_{i\in s} o_i$, $V\left(\hat{Y}\right)$ is defined by,

$$V(\hat{Y}) \approx E_R V_S\left(\sum_{i\in s} o_i \middle| \boldsymbol{R}\right) \approx n\sum_{i\in U} E_R(o_i^2)P_i - \frac{1}{n}\left(\sum_{i\in U} E_R(o_i)P_i\right)^2, \tag{11}$$

where $o_i=f(y_i)$, $P_i=\dfrac{k_i}{\sum_{i\in U} k_i}$, k_i is the size variable, $\boldsymbol{R}=(r_1,r_2,\ldots,r_N)'$ and $E_R V_S(\cdot)$ is the expectation of variance operator respect to nonresponse mechanism and sampling design respectively.

Lawson (2017) proposed a new variance estimator for estimating the population total, $V(\hat{Y})$ is defined as follows.

$$\hat{V}(\hat{Y}) \approx \frac{n}{n-1}\left[\sum_{i\in s}\hat{o}_i^2 - \frac{1}{n}\left(\sum_{i\in s}\hat{o}_i\right)^2\right], \tag{12}$$

where $\hat{o}_i$ is the estimator of o_i when o_i contains an unknown parameter and where otherwise $\hat{o}_i=o_i$.

The Ponkaew & Lawson (2019) variance estimator requires π_{ij} in the estimation. We can see that the Lawson's (2017) estimator in equation (12) does not require π_{ij}. In this section, we propose a new variance estimator following Lawson (2017) using Ponkaew & Lawson's (2019) estimator under PPSWOR when the sampling fraction is negligible. The new variance estimator does not require the joint inclusion. The details are shown as follows.

From the Ponkaew & Lawson (2019) estimator in equation (8) we can write $\hat{Y}_r^{(1)}$ in the following form.

 doi: 10.14416/j.appsci.2019.07.001

$$\hat{Y}_r^{(1)} = \sum_{i\in s} \frac{r_i y_i}{\pi_i p} = \sum_{i\in s} o_i \ , \tag{13}$$

where $o_i = f(y_i) = \frac{r_i y_i}{\pi_i p}$.

The variance of $\hat{Y}_r^{(1)}$, $V(\hat{Y}_r^{(1)})$ is defined by,

$$V(\hat{Y}_r^{(1)}) \approx E_R V_S \left(\hat{Y}_r^{(1)} \middle| \boldsymbol{R} \right) = E_R V_S \left(\sum_{i\in s} o_i \middle| \boldsymbol{R} \right). \tag{14}$$

Following Lawson (2017), $V(\hat{Y}_r^{(1)})$ in equation (14) is given by,

$$V(\hat{Y}_r^{(1)}) \approx n \sum_{i\in U} E_R(o_i^2) P_i - \frac{1}{n} \left(\sum_{i\in U} E_R(o_i) P_i \right)^2 \ , \tag{15}$$

where $o_i = f(y_i) = \frac{r_i y_i}{\pi_i p}$ and $P_i = \frac{k_i}{\sum_{i\in U} k_i}$.

Consider $E_R(o_i)$ and $E_R(o_i^2)$ in equation (15),

$$E_R(o_i) = E_R\left(\frac{r_i y_i}{\pi_i p} \right) = \frac{y_i E_R(r_i)}{\pi_i p} = \frac{y_i p}{\pi_i p} = \frac{y_i}{\pi_i} \ . \tag{16}$$

Let $o_i^2 = \frac{r_i y_i^2}{\pi_i^2 p^2}$, $r_i^2 = r_i$. Then

$$E_R(o_i^2) = E_R\left(\frac{r_i y_i^2}{\pi_i^2 p^2} \right) = \frac{y_i^2}{\pi_i^2 p} \ . \tag{17}$$

Substitute $E_R(o_i)$ in (16) and $E_R(o_i^2)$ in (17) into (15), a new variance estimator is given by,

$$V(\hat{Y}_r^{(1)}) \approx n \sum_{i\in U} \frac{y_i^2}{p\pi_i^2} P_i - \frac{1}{n} \left(\sum_{i\in U} \frac{y_i}{\pi_i} P_i \right)^2 = \frac{n}{p} \sum_{i\in U} \breve{y}_i^2 P_i - \frac{1}{n} \left(\sum_{i\in U} \breve{y}_i P_i \right)^2 , \tag{18}$$

where $\breve{y}_i = \frac{y_i}{\pi_i}$.

The estimator of $V(\hat{Y}_r^{(1)})$ is shown in (19).

$$\hat{V}(\hat{Y}_r^{(1)}) \approx \frac{n}{n-1} \left[\sum_{i\in s} \hat{o}_i^2 - \frac{1}{n} \left(\sum_{i\in s} \hat{o}_i \right)^2 \right] , \tag{19}$$

where $\hat{o}_i = o_i = \frac{r_i y_i}{\pi_i p}$ when p is known and $\hat{o}_i = \frac{r_i y_i}{\pi_i \hat{p}}$ when p is unknown. Then we are able to estimate p by $\hat{p} = \frac{r}{n}$ or $\hat{p} = \sum_{i\in s} \frac{r_i}{\pi_i} \Big/ \sum_{i\in s} \frac{1}{\pi_i}$ (Shao & Steel, 1999).

We can omit the term n/(n-1) in equation (19) when it is closer to 1, then we can write $\hat{V}(\hat{Y}_r^{(1)})$ in the following form.

 doi: 10.14416/j.appsci.2019.07.001

$$\hat{V}(\hat{Y}_r^{(1)}) \approx \sum_{i\in s}\hat{o}_i^2 - \frac{1}{n}\left(\sum_{i\in s}\hat{o}_i\right)^2. \quad (20)$$

We can write the proposed variance estimator in Theorem 1 as follows.

Theorem 1 Suppose that a sample s of size n is selected with probability proportional to size sampling without replacement (PPSWOR) from a finite population $U=\{1,2,\ldots,N\}$ of size N. Let $\hat{Y}_r^{(1)}$ denote the population total estimator under the reverse framework when the response probabilities are uniform and sampling fraction is negligible.

(1) Assume that p is known, the estimated variance of $\hat{Y}_r^{(1)}$ is defined by,

$$\hat{V}_1(\hat{Y}_r^{(1)}) \approx \frac{n}{n-1}\frac{1}{p^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right],$$

$$\hat{V}_2(\hat{Y}_r^{(1)}) \approx \frac{1}{p^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right], \text{ when } \frac{n}{n-1}\approx 1,$$

where $\breve{y}_i = \dfrac{y_i}{\pi_i}$.

(2) Assume that p is unknown, the estimated variance of $\hat{Y}_r^{(1)}$ is defined by,

$$\hat{V}_3(\hat{Y}_r^{(1)}) \approx \frac{n}{n-1}\frac{1}{\hat{p}^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right], \ \hat{p}=\frac{r}{n},$$

$$\hat{V}_4(\hat{Y}_r^{(1)}) \approx \frac{1}{\hat{p}^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right], \ \hat{p}=\frac{r}{n} \text{ and } \frac{n}{n-1}\approx 1.$$

$$\hat{V}_5(\hat{Y}_r^{(1)}) \approx \frac{n}{n-1}\frac{1}{\hat{p}^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right], \ \hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\Big/\sum_{i\in s}\frac{1}{\pi_i},$$

$$\hat{V}_6(\hat{Y}_r^{(1)}) \approx \frac{1}{\hat{p}^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right], \ \hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\Big/\sum_{i\in s}\frac{1}{\pi_i} \text{ and } \frac{n}{n-1}\approx 1,$$

where $\breve{y}_i = \dfrac{y_i}{\pi_i}$.

Proof

(1) Assume that p is known,

Substitute $\hat{o}_i = o_i = \dfrac{r_i y_i}{\pi_i p}$ into (19), we can write $\hat{V}(\hat{Y}_r^{(1)})$ with $\hat{V}_1(\hat{Y}_r^{(1)})$

$$\begin{aligned}\hat{V}_1(\hat{Y}_r^{(1)}) &\approx \frac{n}{n-1}\left[\sum_{i\in s}\left(\frac{r_i y_i}{\pi_i p}\right)^2 - \frac{1}{n}\left(\sum_{i\in s}\frac{r_i y_i}{\pi_i p}\right)^2\right] \\ &= \frac{n}{n-1}\frac{1}{p^2}\left[\sum_{i\in s}\frac{r_i y_i^2}{\pi_i^2} - \frac{1}{n}\left(\sum_{i\in s}\frac{r_i y_i}{\pi_i}\right)^2\right] \\ &= \frac{n}{n-1}\frac{1}{p^2}\left[\sum_{i\in s} r_i\breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i\breve{y}_i\right)^2\right].\end{aligned}$$

 doi: 10.14416/j.appsci.2019.07.001

Then

$$\hat{V}_1(\hat{Y}_r^{(1)}) \approx \frac{n}{n-1}\frac{1}{p^2}\left[\sum_{i\in s} r_i \breve{y}_i^2 - \frac{1}{n}\left(\sum_{i\in s} r_i \breve{y}_i\right)^2\right], \tag{21}$$

where $\breve{y}_i = \dfrac{y_i}{\pi_i}$. We can find $\hat{V}_2(\hat{Y}_r^{(1)})$ by replacing $\dfrac{n}{n-1}$ in (21) with 1.

(2) Assume that p is unknown, the proof is similar to (1).

Note that when $n>2$ then $n>n-1$ and $\dfrac{n}{n-1}>1$. When p is known then $\hat{V}_2(\hat{Y}_r^{(1)}) < \hat{V}_1(\hat{Y}_r^{(1)})$ and when p is unknown then $\hat{V}_4(\hat{Y}_r^{(1)}) < \hat{V}_3(\hat{Y}_r^{(1)})$ and $\hat{V}_6(\hat{Y}_r^{(1)}) < \hat{V}_5(\hat{Y}_r^{(1)})$.

Results and discussion

A simulation study has been conducted to see the performance of the new variance estimator compared to the Ponkaew & Lawson (2019) estimator. In both situations; p is known and p is unknown, however Ponkaew & Lawson's (2019) estimator requires π_{ij} which is usually unknown in unequal probability sampling therefore we can apply the Hájek (1981) method to estimate unknown π_{ij} when π_i is known for all $i \in U$, the estimated value of π_{ij} is given by,

$$\hat{\pi}_{ij} \approx \pi_i \pi_j [1-(1-\pi_i)(1-\pi_j)d^{-1}], \tag{22}$$

where $d = \sum_{i\in U} \pi_i(1-\pi_i)$.

Substitute $\hat{\pi}_{ij}$ from (22) into (9) and (10), then the Ponkaew & Lawson (2019) estimator is given by,

$$\hat{V}_{00}(\hat{Y}_r^{(1)}) = \frac{1}{p^2}\left[\sum_{i\in s}\frac{(1-\pi_i)}{\pi_i^2} r_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{\hat{D}}_{ij} r_i r_j y_i y_j\right], \text{ when } p \text{ is known }, \tag{23}$$

$$\hat{V}_{01}(\hat{Y}_r^{(1)}) = \left(\frac{\sum_{i\in s}\frac{1}{\pi_i}}{\sum_{i\in s}\frac{r_i}{\pi_i}}\right)^2 \left[\sum_{i\in s}\frac{(1-\pi_i)}{\pi_i^2} r_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{\hat{D}}_{ij} r_i r_j y_i y_j\right], \text{ when } p \text{ is unknown }, \tag{24}$$

where $\hat{\hat{D}}_{ij} = \dfrac{\hat{\pi}_{ij} - \pi_i\pi_j}{\hat{\pi}_{ij}\pi_i\pi_j}$.

We compare the efficiency of the new estimator with the Ponkaew & Lawson (2019) estimator via simulation study using the relative bias as a criterion following Thompson et al. (2010) where $y_i = k_i'\boldsymbol{\beta} + \varepsilon_i$. The simulation steps are shown as below.

Step 1 In order to generate y_i, we generate $k_i \sim Gam(4,5)$, $\varepsilon_i \sim N(0,0.7)$ and $\beta_0 = 0$, $\beta_1 = 1$, $i = 1,2,\ldots,N$ where $N = 10{,}000$.

Step 2 A samples s of size $n = 25,\ 30,\ 50,\ 100,\ 300$ and 500 is selected using Midzuno (1952) scheme.

 doi: 10.14416/j.appsci.2019.07.001

Step 3 In each sampling and sample size n, nonresponse is generated using a uniform response mechanism with probabilities of response $p = 0.7$.

Step 4 Compute the new variance estimators and the existing estimators.

Step 5 Repeat steps (2) to (4) for 5,000 times (M=5,000).

Step 6 The simulation relative bias of the variance estimator $\hat{V}(Y_r^{(m)})$ is computed,

$RB_{SIM}[\hat{V}(\hat{Y}_r^{(m)})] = \dfrac{E_{SIM}[\hat{V}(\hat{Y}_r^{(m)})] - V_{SIM}(\hat{Y}_r^{(m)})}{V_{SIM}(\hat{Y}_r^{(m)})}$ where $E_{SIM}[\hat{V}(\hat{Y}_r^{(m)})] = \dfrac{1}{M}\sum_{j=1}^{M}\hat{V}\left[\hat{Y}_{r(j)}^{(m)}\right]$ is the simulated expectation of the variance estimator and $V_{SIM}(\hat{Y}_r^{(m)}) = \dfrac{1}{M-1}\sum_{j=1}^{M}\left[\hat{Y}_{r(j)}^{(m)} - E_{SIM}(\hat{Y}_r^{(m)})\right]^2$ is the simulated variance of the estimator $\hat{Y}_r^{(m)}$, $m=1,2$. $E_{SIM}(\hat{Y}_r^{(m)})$ is defined by $E_{SIM}(\hat{Y}_r^{(m)}) = \dfrac{1}{M}\sum_{j=1}^{M}\hat{Y}_{r(j)}^{(m)}$.

The simulation relative bias for the new estimator and all other existing estimators are shown in Table 1 where p is known and in Table 2 where p is unknown.

Table 1. The simulation relative bias (%) of $\hat{V}(Y_r^{(m)})$ when p is known

n	Simulation Relative Bias (%)		
	$\hat{V}_{00}(\hat{Y}_r^{(1)})$	$\hat{V}_1(\hat{Y}_r^{(1)})$	$\hat{V}_2(\hat{Y}_r^{(1)})$
25	18.56	18.76	**18.27**
30	17.65	17.73	**16.29**
50	16.36	**16.14**	**15.72**
100	6.82	**5.34**	**3.23**
300	6.53	**6.46**	**6.02**
500	6.38	**5.42**	**4.37**

From Table 1 we can see that the new variance estimator, $\hat{V}_2(\hat{Y}_r^{(1)})$ when $n/(n-1)$ is omitted performs well when compared to the Ponkaew & Lawson(2019) estimator, $\hat{V}_{00}(\hat{Y}_r^{(1)})$ when p is known. The $\hat{V}_1(\hat{Y}_r^{(1)})$ also gives smaller percent relative bias when compared to $\hat{V}_{00}(\hat{Y}_r^{(1)})$ except for n=25 and n=30.

We can also see similar results in Table 2 when p is unknown. All new variance estimators perform well when compared to the Ponkaew & Lawson (2019) estimator when $n/(n-1)$ is omitted. Some new variance estimators only give slightly higher percent relative bias when compared to the exisiting estimator for n=25 and n=30. The variance estimators when p is unknown give higher simulation percent relative bias when compared to the variance estimators when p is known. Therefore, it is an alternative useful variance estimator with free joint inclusion probability.

Conclusions

An alternative variance estimator with free joint inclusion probability for estimating the population total has been proposed under unequal probability sampling without replacement in the presence of nonresponse. We considered it under the reverse framework where the sampling fraction is negligible and the response probabilities are uniform. A simulation study was employed to see the performance of the new variance estimators compared to the existing

doi: 10.14416/j.appsci.2019.07.001

estimators in both cases where p is known and where p is unknown. We can see that the new variance estimators performed well, especially when $n/(n-1)$ is omitted and only gives slightly higher percent relative bias when compared to the existing estimators when $n/(n-1)$ is not closed to one.

Table 2. The simulation relative bias (%) of $\hat{V}(Y_r^{(m)})$ when p is unknown

n	Simulation Relative Bias (%)				
	$\hat{V}_{01}(\hat{Y}_r^{(1)})$	$\hat{V}_3(\hat{Y}_r^{(1)})$	$\hat{V}_4(\hat{Y}_r^{(1)})$	$\hat{V}_5(\hat{Y}_r^{(1)})$	$\hat{V}_6(\hat{Y}_r^{(1)})$
25	225.85	232.30	227.01	226.39	**217.34**
30	201.56	208.38	**196.04**	202.29	**190.20**
50	159.24	**157.53**	**156.67**	**156.93**	**156.08**
100	130.50	**123.33**	**118.87**	**121.59**	**117.16**
300	105.52	**101.83**	**101.16**	**101.55**	**100.88**
500	104.61	**97.91**	**95.93**	**96.56**	**94.60**

Acknowledgements

This work was funded by The Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand. Contract no 6245103. We appreciate the unknown referees for their helpful comments and we would like to thank them for helping to improve our paper.

References

Berger, Y. G. (1998). Rate of convergence for asymptotic variance for the Horvitz-Thompson estimator. *Journal of Statistical Planning and Inference*, *74*(1), 149-168. doi: 10.1016/S0378-3758(98)00107-4

Berger, Y. G. (2003). A modified Hájek variance estimator for systematic sampling. *Statistics in Transition*, *6*(1), 5-21.

Hájek, J. (1981). *Sampling from a Finite Population*. New York: Marcel Dekker Inc.

Horvitz, D. G. & Thompson D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, *47*(260), 663-685. https://doi.org/10.1080/01621459.1952.10483446

Lawson, N. (2017). *Variance estimation in the presence of nonresponse under probability proportional to size sampling.* pp. 116-119. In the 6th Annual International Conference on Computational Mathematics, Computational Geometry & Statistics (CMCGS 2017) and 5th Annual International Conference on Operations Research and Statistics (ORS 2017), 6-7 March 2017, Hotel Fort Canning, Singapore. doi: 10.5176/2251-1911_CMCGS17.32

Midzuno, H. (1952). On sampling system with probability proportional to sum of sizes. *Annals of the Institute of Statistical Mathematics*, *3*, 99-107.

Ponkaew, C. & Lawson, N. (2019). Estimating variance in the presence of nonresponse under unequal probability sampling. *Suranaree Journal of Science and Technology*. (In press)

Särndal, C. E. & Lundström, S. (2005). *Estimation in Surveys with Nonresponse* (1st ed.). Chichester, England: John Wiley & Sons.

Shao, J. & Steel, P. (1999). Variance estimation for survey data with composite imputation and nonnegligible sampling fractions. *Journal of the American Statistical Association, 94*(445), 254-265. doi: 10.1080/01621459.1999.10473841

Thompson, K. J., Yung, W. & Haziza, D. (2010). The effect of nonresponse adjustments on variance estimation. *Survey Methodology*, *36*(1), 35-43.