

Comparison between Hansen-Hurwitz and Horvitz-Thompson estimators for adaptive cluster sampling


MOHAMMAD SALEHI M.

School of Mathematical Sciences, Isfahan University of Technology, Isfahan, 84156, Iran

Received August 2000; Revised January 2002

Thompson (1990) introduced the adaptive cluster sampling design. This sampling design has been shown to be a useful sampling method for parameter estimation of a clustered and scattered population (Roesch, 1993; Smith *et al.*, 1995; Thompson and Seber, 1996). Two estimators, the modified Hansen-Hurwitz (HH) and Horvitz-Thompson (HT) estimators, are available to estimate the mean or total of a population. Empirical results from previous researches indicate that the modified HT estimator has smaller variance than the modified HH estimator. We analytically compare the properties of these two estimators. Some results are obtained in favor of the modified HT estimator so that practitioners are strongly recommended to use the HT estimator despite easiness of computations for the HH estimator.

Keywords: adaptive sampling, network sampling, Rao-Blackwell theorem

1352-8505 © 2003  Kluwer Academic Publishers

1. Introduction

1.1 *Adaptive cluster sampling*

During the last decade, adaptive cluster sampling designs have been combined with conventional sampling designs and used in different research fields. Adaptive cluster sampling (ACS) has been extended to the use of primary units (Thompson, 1991a), stratified ACS (Thompson, 1991b), restricted ACS (Brown, 1994; Brown and Manly, 1998; Salehi and Seber, 2002), ACS based on order statistics (Thompson, 1996), ACS with networks selected without replacement (Salehi and Seber, 1997a), and two-stage ACS (Salehi and Seber, 1997b). Quinn II *et al.* (1999) used a restricted adaptive cluster sampling in a fisheries survey, Roesch (1993) used unequal probability adaptive cluster sampling in forestry and researchers in other fields applied adaptive cluster samplings, many of those applications were in biology (Smith *et al.*, 1995; Lo *et al.*, 1997). Two estimators, Horvitz-Thompson and Hansen-Hurwitz estimators were modified by Thompson (1990) for adaptive cluster sampling. These two estimators were also developed for new extensions of adaptive cluster sampling (e.g., stratified ACS, two stage ACS etc.). The modified HH estimators are easier to calculate so that practitioner have

tended to use them more often; however, the modified HT estimators have had smaller variances in previous applications. In this article, we compare properties of these two estimators and give some more concrete results in favor of the modified HT estimator. The results can be easily extended to different adaptive cluster sampling designs recently developed.

In this section, we briefly describe the adaptive cluster sampling design. Suppose that we have a population of units. An initial sample of units is selected by some conventional sampling design. Whenever the value of the variable of interest (or any associated variable) of a selected unit satisfies a specified condition, say C , its neighboring units are added to the sample. Furthermore, if any other units in these neighboring units satisfy C then their neighborhoods are also added to the sample. This process continues until a cluster of units is formed with a boundary of units, called edge units, which do not satisfy C . A cluster without its edge units forms a network, and a unit not satisfying the condition C also forms a network of size one. The networks are disjoint and form a partition of population units. The condition for extra sampling is defined on the value of a variable which is known only for units in the sample.

1.2 Notation and estimators

Consider a population of N units (u_1, u_2, \dots, u_N) labeled by $(1, 2, \dots, N)$. With u_i is associated a variable of interest y_i , for $i = 1, 2, \dots, N$. A simple random sample of size n_1 is taken without replacement. Further units are then added adaptively using condition C . Following the notation of Thompson and Seber (1996), suppose that the final ordered sample of the labels is $s_o = (i_1, i_2, \dots, i_n)$. We note that repeated y -values can occur in this sample. Let d_o denote the data vector whose components are the unit labels and their corresponding y -values, namely $d_o = ((i, y_i) : i \in s_o)$. Consider the unordered reduced set $s_R = \{i_1, i_2, \dots, i_\nu\}$ of the ν distinct labels in s_o . Then we have $d_R = \{(i, y_i) : i \in s_R\}$ is minimal sufficient for d_o in any adaptive sampling design (Basu, 1969; Thompson and Seber, 1996).

Let A_i denote the network containing unit i and let α_i denote the probability that the initial sample intersects A_i ,

$$\alpha_i = 1 - \frac{\binom{N-m_i}{n_1}}{\binom{N}{n_1}}, \quad (1)$$

where m_i is the number of units in A_i . The modified HT estimator for the mean is

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{k=1}^{\kappa} \frac{y_k^*}{\alpha_k}, \quad (2)$$

where y_k^* is the sum of the y -values for the k th network, κ is the number of distinct networks intersected by the initial sample. We have $\alpha_k = \alpha_i$ for every unit i in network k . The joint inclusion probability of networks j and k is given by

$$\alpha_{jk} = 1 - \frac{\left[\binom{N-m_j}{n_1} + \binom{N-m_k}{n_1} - \binom{N-m_j-m_k}{n_1} \right]}{\binom{N}{n_1}} \quad (3)$$

and the variance of $\hat{\mu}_{HT}$ is

$$\text{var}[\hat{\mu}_{HT}] = \frac{1}{N^2} \left[\sum_{j=1}^K \sum_{k=1}^K y_j^* y_k^* \left(\frac{\alpha_{jk} - \alpha_j \alpha_k}{\alpha_j \alpha_k} \right) \right], \quad (4)$$

where K denotes the number of networks in the population. An unbiased estimator of the above variance is

$$\widehat{\text{var}}[\hat{\mu}_{HT}] = \frac{1}{N^2} \left[\sum_{j=1}^K \sum_{k=1}^K \frac{y_j^* y_k^*}{\alpha_{jk}} \left(\frac{\alpha_{jk}}{\alpha_j \alpha_k} - 1 \right) \right], \quad (5)$$

where α_{jj} is interpreted as α_j . The variance estimator of Horvitz-Thompson estimator can take negative value for some conventional sampling designs. Salehi and Seber (1997b) prove that (5) is non-negative for $n_1 = 2$.

The modified HH estimator for the mean is

$$\hat{\mu}_{HH} = \frac{1}{n_1} \sum_{i=1}^{n_1} w_i, \quad (6)$$

where w_i is the mean of the m_i observations in A_i . Here $\hat{\mu}_{HH}$ can be recognized as the sample mean obtained by taking a simple random sample of size n_1 from a population of w_i values (Thompson and Seber, 1996). Using the theory of simple random sampling we have

$$\text{var}[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1 (N - 1)} \sum_{i=1}^N (w_i - \mu)^2, \quad (7)$$

with unbiased estimate

$$\widehat{\text{var}}[\hat{\mu}_{HH}] = \frac{N - n_1}{N n_1 (n_1 - 1)} \sum_{i=1}^{n_1} (w_i - \hat{\mu}_{HH})^2. \quad (8)$$

If the initial sample is selected with replacement, both estimators can be obtained in a similar fashion (Thompson and Seber, 1996, page 100).

2. Mathematical comparison

Both variances are unchanged by within-network variation because they involve the sum of the y -values over networks. Since within-network variation does affect the variance of the sample mean from a simple random sample (SRS), the relative efficiency of these adaptive estimators to SRS will increase as within-network variance increases. The modified HH can be written as,

$$\hat{\mu}_{HH} = \frac{1}{Nn_1} \sum_{i=1}^{n_1} \frac{y_i^*}{p_i},$$

where $p_i = m_i/N$ is the probability that network A_i is selected in a draw. We note that if the y_i^* s are proportional to p_i then $\text{var}[\hat{\mu}_{HH}]$ would be zero. If the y_k^* s are proportional to α_k in (2) then $\text{var}[\hat{\mu}_{HT}]$ would be very close to zero. Hence, it is sensible to say that the estimator for which the appropriate proportionality property holds more closely is the more efficient estimator (Särndal *et al.*, 1992 page 88). A typical condition for extra sampling is of the form $y_i > c$ for some constant c . When the population is highly clumped, the networks of size one which satisfy the condition $y_i > c$ should be very few so that almost all of the networks of size one do not satisfy the condition. Hence, we have a reasonably big gap between y_k^* of the networks of size one and y_k^* of the networks of size two and greater. We now investigate which of the probabilities α_k and p_i show this similar uneven behavior. Suppose α_k is a function of m_k . Looking at the first order difference of α_k , namely

$$\Delta\alpha_k(m_k) = \alpha_k(m_k + 1) - \alpha_k(m_k) = \frac{\binom{N-m_k-1}{n_1}}{\binom{N}{n_1}} \frac{n_1}{N - m_k - n_1} \geq 0,$$

we see that $\alpha_k(m_k)$ is an increasing function, as intuitively expected. We point out that $\Delta\alpha_k(m_k) = 0$ when $N \leq m_k + n_1$. The second order difference of $\alpha_k(m_k)$ is,

$$\Delta^2\alpha_k(m_k) = \frac{\binom{N-m_k-2}{n_1}}{\binom{N}{n_1}} \frac{n_1 - n_1^2}{(N - m_k - n_1)(N - m_k - n_1 - 1)},$$

which is non-positive for $n_1 \geq 1$. Therefore, the graph (polygon) of $\alpha_k(m_k)$ is concave downwards which implies that the rate of growth decreases as m_k increases. In other words, $\Delta\alpha_k(1)$ is the maximum of $\Delta\alpha_k(m_k)$ so that the biggest jump in the value of $\alpha_k(m_k)$ occurs when m_k moves from 1 to 2. However, p_i is a linear function of m_i and does not have such a jump. Therefore, we can expect that $\hat{\mu}_{HT}$ performs better than $\hat{\mu}_{HH}$ because it more closely approximates the proportionality property. In Fig. 1, the graphs of $\alpha_k(m_k)$ are drawn for different initial samples when $N = 400$. As $\Delta\alpha_k(1)$ increases with n_1 , the jump of $\alpha_k(m_k)$ from 1 to 2 also increases with n_1 . Hence if the y -values of the units satisfying condition C are relatively bigger than those not satisfying condition C , the relative efficiency of $\hat{\mu}_{HT}$ with respect to $\hat{\mu}_{HH}$ is much greater for larger initial samples. This fact on relative efficiency is consistent with Thompson (1990), Roesch (1993), and Smith *et al.* (1995).

3. Statistical comparison

We introduce statistic d_J , which consists of d_R and indicator functions J_i . The indicator variable J_i equals 1 when the initial sample intersects the network that contains u_i and 0 otherwise. Thus,

$$d_J = \{(i, y_i, J_i = j) : i \in s_R\}.$$

The minimal sufficient statistic d_R can be expressed as a function of d_J so that the statistic

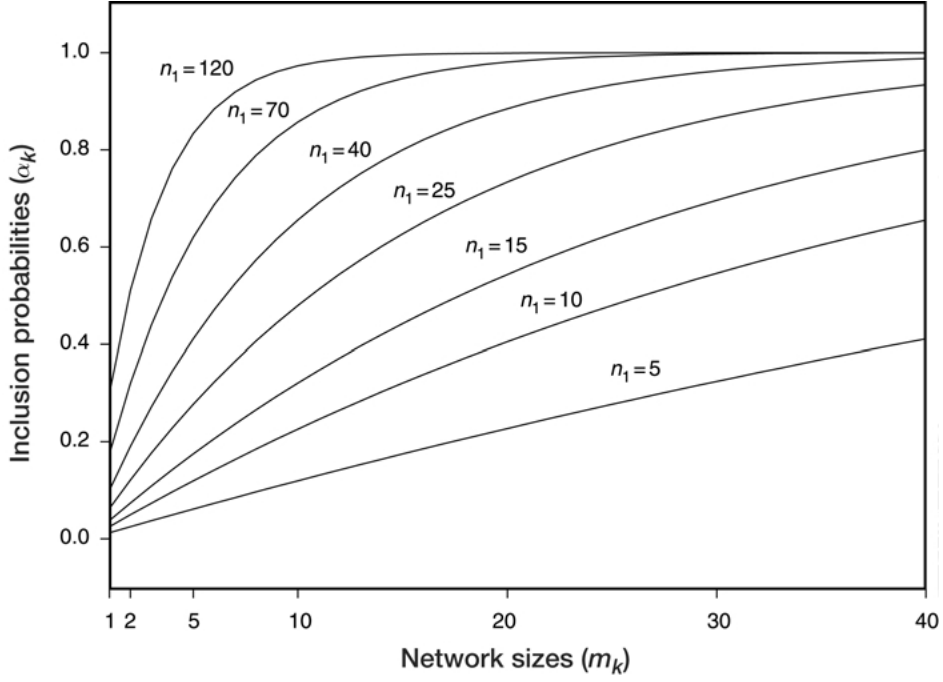


Figure 1. Inclusion probability functions in terms of network size for initial sample size (5, 10, 15, 25, 40, 70, 120) where $N = 400$.

d_J is sufficient. Hence $E(\hat{\mu} \cdot | d_J)$ does not depend on unknown parameters and can therefore be an estimator (\cdot stands for HH or HT). We now show that if we condition on d_J , as a reduction of the data set, we gain no improvement for $\hat{\mu}_{HT}$, but $\hat{\mu}_{HH}$ can be improved. This means that $\hat{\mu}_{HT}$ has already used all information in d_J , but $\hat{\mu}_{HH}$ has not. The modified Horvitz-Thompson estimator can be rewritten,

$$\hat{\mu}_{HT} = \frac{1}{N} \sum_{i \in s_R} \frac{y_i}{\alpha'_i} J_i.$$

Let s_R be made up of s_e adaptively added edge units and s_1 remaining units. Then $J_i = 0$ for $i \in s_e$ and $J_i = 1$ for $i \in s_1$. Recalling that κ is the number of distinct networks in the sample,

$$\begin{aligned} E[\hat{\mu}_{HT} | d_J] &= \frac{1}{N} \sum_{i \in s_R} \frac{y_i}{\alpha'_i} E[J_i | d_J] \\ &= \frac{1}{N} \sum_{i \in s_1} \frac{y_i}{\alpha'_i} \\ &= \frac{1}{N} \sum_{k=1}^{\kappa} \frac{y_k^*}{\alpha_k} \\ &= \hat{\mu}_{HT}, \end{aligned}$$

where $\alpha_k = \alpha'_i$ for any unit i in network k (we do not distinguish between the random variable $\hat{\mu}_{\text{HT}}$ and its value for notational convenience). Therefore, $\hat{\mu}_{\text{HT}}$ has not improved using d_J .

We now show that $\hat{\mu}_{\text{HH}}$ will be improved using d_J . Let I_i be an indicator function which takes the value 1 when unit i is selected in the initial sample and 0 otherwise. From (6) we have,

$$\begin{aligned}\tilde{\mu}_{\text{HH}} &= E[\hat{\mu}_{\text{HH}}|d_J] = \frac{1}{n_1} \sum_{i=1}^N w_i E[I_i|d_J] \\ &= \frac{1}{n_1} \sum_{i=1}^N w_i \Pr(I_i = 1|d_J).\end{aligned}$$

Let ξ and ξ_i be respectively the number of possible initial samples that give rise to d_J and the number of initial samples containing unit i compatible with d_J . Since $\Pr(I_i = 1|d_J) = \xi_i/\xi$ when unit i is in one of the networks intersected by the initial sample and $\Pr(I_i = 1|d_J) = 0$ otherwise. Thus, we have

$$\tilde{\mu}_{\text{HH}} = \frac{1}{n_1 \xi} \sum_{i=1}^m w_i \xi_i. \quad (9)$$

We now find ξ and ξ_i . The number of units with $J_i = 1$ in d_J is $m. = \sum_{k=1}^{\kappa} m_k$. We should choose n_1 from the $m.$ units. This can be done in $\binom{m.}{n_1}$ ways but not all of them can give rise to d_J (i.e., are compatible with d_J). Those combinations that do not include at least one unit from each of the κ distinct networks intersected by the initial sample are not compatible with d_J . Suppose C_k is the set of combinations that contains no unit from network k . The set $\bigcup_{k=1}^{\kappa} C_k$ is the set of possible samples that contains no units from at least one of those κ networks. Thus, ξ equals $\binom{m.}{n_1} - |\bigcup_{k=1}^{\kappa} C_k|$, where $||$ stands for cardinality. This is given by

$$\xi = \binom{m.}{n_1} - \sum_{g=1}^{\kappa} \binom{m. - m_g}{n_1} + \sum_{g,h} \binom{m. - m_g - m_h}{n_1} + \dots + (-1)^{\kappa} \binom{m. - \sum m_g}{n_1}.$$

The last term of ξ is zero but we leave it in for notational convenience. If unit i forms a network of size one, say network k , then all ξ possible samples contain network k . If the size of network k is greater than one, then the number of combinations of n_1 from the $m.$ units which contain a particular unit i in the network k is $\binom{m.-1}{n_1-1}$. Since we know that there is at least one unit from network k in those combinations, we just exclude the combinations which do not contain at least one unit from each of remaining $\kappa - 1$ networks. Thus, ξ_i is given by

$$\xi_i = \begin{cases} \xi & \text{unit } i \text{ forms a network,} \\ \binom{m.-1}{n_1-1} - \sum_{g \neq k}^{\kappa} \binom{m.-1-m_g}{n_1-1} \\ \quad + \sum_{g,h \neq k} \binom{m.-1-m_g-m_h}{n_1-1} \\ \quad + \dots + (-1)^{\kappa} \binom{m.-1-\sum_{g \neq k} m_g}{n_1-1} & \text{otherwise.} \end{cases}$$

Since the ξ_i s are equal for all m_k units i in network k ($= \eta_k$, say), and w_i is repeated m_k times

$$\tilde{\mu}_{HH} = \frac{1}{n_1 \zeta} \sum_{k=1}^{\kappa} y_k^* \eta_k. \quad (10)$$

Its variance is given by,

$$\text{var}[\tilde{\mu}_{HH}] = \text{var}[\hat{\mu}_{HH}] - E(\text{var}[\hat{\mu}_{HH}|d_J]). \quad (11)$$

An unbiased estimator for the first term of the right hand side of (11) is $\widehat{\text{var}}[\hat{\mu}_{HH}]$, given by (8). An unbiased estimator for the second term of the right hand side of (11) is $\text{var}[\hat{\mu}_{HH}|d_J]$. To evaluate $\text{var}[\hat{\mu}_{HH}|d_J]$, we need to know the number of possible initial samples containing two units i and j , say ξ_{ij} . If both units i and j are from networks of size one, then all possible samples contain them and $\xi_{ij} = \zeta$. If unit j forms a network of size one and unit i does not, then units j is in all possible samples and ξ_{ij} can be evaluated like ζ_i . If units i and j are both in a network, say network k , we consider all the combinations containing them and then exclude those combinations not containing at least one unit from each of the $\kappa - 1$ remaining networks. If unit i and j are in different networks with sizes greater than one, we consider all the combinations containing them and then exclude those combinations not containing at least one unit from each of the $\kappa - 2$ remaining networks. Hence

$$\xi_{ij} = \begin{cases} \zeta & i, j \text{ are networks} \\ & \text{of sizes 1,} \\ & j \text{ is a network} \\ & \text{of size one, } i \in k, \\ \left(\binom{m-1}{n_1-1} - \sum_{g \neq k}^{\kappa} \binom{m-m_g-1}{n_1-1} + \sum_{g, h \neq k} \binom{m-m_g-m_h-1}{n_1-1} \right) & \\ + \dots (-1)^{\kappa} \binom{m-\sum m_g-1}{n_1-1} & \\ \left(\binom{m-2}{n_1-2} - \sum_{g \neq k}^{\kappa} \binom{m-m_g-2}{n_1-2} + \sum_{g, h \neq k} \binom{m-m_g-m_h-2}{n_1-2} \right) & i, j \in k, \\ + \dots (-1)^{\kappa} \binom{m-\sum m_g-2}{n_1-2} & \\ \left(\binom{m-2}{n_1-2} - \sum_{g \neq k, l}^{\kappa} \binom{m-m_g-2}{n_1-2} + \sum_{g, h \neq k, l} \binom{m-m_g-m_h-2}{n_1-2} \right) & i \in k, j \in l \\ + \dots (-1)^{\kappa} \binom{m-\sum m_g-2}{n_1-2} & k \neq l, \end{cases}$$

and $Pr(I_i = 1, I_j = 1|d_J) = \xi_{ij}/\zeta$ where i and j are in networks intersected by the initial sample. Now, $\text{var}[\hat{\mu}_{HH}|d_J]$ is given by

$$\begin{aligned} \text{var}[\hat{\mu}_{HH}|d_J] &= \frac{1}{n_1^2} \text{var} \left[\sum_{i=1}^N w_i I_i | d_J \right] \\ &= \frac{1}{n_1^2} \left(\sum_{i=1}^N w_i^2 \text{var}[I_i | d_J] + 2 \sum_{i=1}^N \sum_{j < i} w_i w_j \text{cov}[I_i, I_j | d_J] \right) \\ &= \frac{1}{n_1^2} \left(\sum_{i=1}^N w_i^2 [Pr(I_i = 1|d_J) - Pr(I_i = 1|d_J)^2] \right) \end{aligned}$$

$$\begin{aligned}
& + 2 \sum_{i=1}^N \sum_{j < i} w_i w_j [Pr(I_i = 1, I_j = 1 | d_J) \\
& - Pr(I_i = 1 | d_J) Pr(I_j = 1 | d_J)] \Bigg) \\
& = \frac{1}{n_1^2 \xi^2} \left(\sum_{i=1}^{m.} w_i^2 (\xi \xi_i - \xi_i^2) + 2 \sum_{i=1}^{m.} \sum_{j < i} w_i w_j (\xi_{ij} \xi - \xi_i \xi_j) \right). \quad (12)
\end{aligned}$$

We should point out that the $\widehat{\text{var}}[\hat{\mu}_{HH}]$ can be improved using Rao-Blackwell theorem, namely

$$\begin{aligned}
\widehat{\text{var}}_{\text{RB}}[\hat{\mu}_{HH}] &= E(\widehat{\text{var}}[\hat{\mu}_{HH}] | d_J) \\
&= \frac{1}{\xi} \sum_{g=1}^{\xi} \widehat{\text{var}}_g[\hat{\mu}_{HH}], \quad (13)
\end{aligned}$$

and using it in (11), we have

$$\begin{aligned}
\widehat{\text{var}}_{\text{RB}}[\tilde{\mu}_{HH}] &= \frac{1}{\xi} \sum_{g=1}^{\xi} \widehat{\text{var}}_g[\hat{\mu}_{HH}] \\
&- \frac{1}{n_1^2 \xi^2} \left(\sum_{i=1}^{m.} w_i^2 (\xi \xi_i - \xi_i^2) + 2 \sum_{i=1}^{m.} \sum_{j < i} w_i w_j (\xi_{ij} \xi - \xi_i \xi_j) \right).
\end{aligned}$$

Discussion in this section has been presented in a statistical inference manner, which can be simply expressed as the following. Ignoring information from the edge units, there is an estimator, $\tilde{\mu}_{HH}$, which uniformly has smaller variance than $\hat{\mu}_{HH}$. This is a disadvantage for $\hat{\mu}_{HH}$ but $\hat{\mu}_{HT}$ does not have such a disadvantage. Thus, a possible competitor for $\hat{\mu}_{HT}$ is $\tilde{\mu}_{HH}$ rather than $\hat{\mu}_{HH}$.

Adaptive cluster sampling was motivated by the problem of sampling a scattered, but clustered population. We therefore expect that the population contains very few networks of size greater than one, and most of the networks of size one correspond to units which do not satisfy condition C. Consequently we can substantially reduce the amount of computation for ξ , ξ_i , and ξ_{ij} by changing the definitions of $m.$ and n_1 in their formulae.

4. Examples

To illustrate the computations associated with $\tilde{\mu}_{HH}$ and clarify the notation, two examples are worked through. In example 2, we use the following algorithm for evaluating $\tilde{\mu}_{HH}$. Since all of the networks of size one with $J_i = 1$ must be intersected by the initial sample, we must allocate an initial sample for every network of size one with J_i s equal to 1. Let κ^* be the number of networks with size greater than 1. To find out the number of possible samples, the remaining initial samples, say n_1^* , must be allocated in the remaining units, say $m.^*$, such that the given samples are compatible with d_J . To achieve that, we should just use $m.^*$, n_1^* and κ^* instead of $m.$, n_1 and κ in the formulae for ξ , ξ_i and ξ_{ij} . This

Table 1. All possible outcomes of adaptive cluster sampling for a population of seven units with y -values 12, 1000, 4, 0, 5, 500, 30 in which the neighborhood of each unit consists of itself plus adjacent units.

<i>Samples</i>	$\hat{\mu}_{HH}$	$\hat{\mu}_{HT}$	$\tilde{\mu}_{HH}$	<i>Samples</i>	$\hat{\mu}_{HH}$	$\hat{\mu}_{HT}$	$\tilde{\mu}_{HH}$
1, 2, 3;	338.7	203.7	338.7	2, 3, 7;1, 5, 7	258.3	309.7	258.3
1, 2, 4;3	337.3	202.4	337.3	2, 4, 5;1, 3	170.3	204.1	170.3
1, 2, 5;3	339.0	204.7	339.0	2, 4, 6;1, 3, 5, 7	257.0	308.4	257.0
1, 2, 6;3, 5, 7	425.7	308.4	385.5	2, 4, 7;1, 3, 5, 6	257.0	308.4	257.0
1, 2, 7;3, 5, 6	425.7	308.4	385.5	2, 5, 6;1, 3, 7	258.7	310.1	258.7
1, 3, 4;2	170.0	203.7	170.0	2, 5, 7;1, 3, 6	258.7	310.1	258.7
1, 3, 5;2	171.7	205.4	171.7	2, 6, 7;1, 3, 5	345.3	308.4	385.5
1, 3, 6;2, 5, 7	258.3	309.7	258.3	3, 4, 5;	3.0	3.0	3.0
1, 3, 7;2, 5, 6	258.3	309.7	258.3	3, 4, 6;5, 7	89.7	107.3	89.7
1, 4, 5;2	170.3	204.1	170.3	3, 4, 7;5, 6	89.7	107.3	89.7
1, 4, 6;2, 3, 5, 7	257.0	308.4	257.0	3, 5, 6;7	91.3	109.0	91.3
1, 4, 7;2, 3, 5, 7	257.0	308.4	257.0	3, 4, 7;5, 6	91.3	109.0	91.3
1, 5, 6;2, 3, 7	258.7	310.1	258.7	3, 6, 7;5	178.0	107.3	178.0
1, 5, 7;2, 3, 6	258.7	310.1	258.7	4, 5, 6;7	90.0	107.7	90.0
1, 6, 7;2, 3, 5	345.3	308.4	385.5	4, 5, 7;6	90.0	107.7	90.0
2, 3, 4;1	170.0	203.7	170.0	4, 6, 7;5	176.7	106.0	176.7
2, 3, 5;1	171.7	205.4	171.7	5, 6, 7;	178.3	107.7	178.3
2, 3, 6;1, 5, 7	258.3	309.7	258.3				
Mean					221.6	221.6	221.6
Variance					9806.5	8286.9	9662.1

An initial sample of three units by simple random sampling without replacement. Whenever an observed y -value exceeds 10, the neighboring units are added to the sample. Initial observed labels are separated from subsequent observed labels in the table by a semicolon. For each possible sample, the value of each estimator is given. The bottom line of the table gives the variance for each estimator.

algorithm is useful when the number of networks of size one in the sample is relatively large.

Example 1. (A small population) In this example the adaptive cluster sampling is applied to a very small population to shed light on the computations and properties of $\tilde{\mu}_{HH}$ in relation to $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$. The population consists of just seven units, with their labels and y -values given by

$$\{(1, 12), (2, 1000), (3, 4), (4, 0), (5, 5), (6, 500), (7, 30)\}.$$

The neighborhood of each unit includes all adjacent units (of which there are either one or two). The condition is defined by $C = \{y : y > 10\}$. The initial sample size is $n_1 = 3$. With the adaptive design in which the initial sample is selected by simple random sampling without replacement, there are $\binom{7}{3} = 35$ possible samples, each having probability $1/35$. The resulting observations and the values of each estimator are listed in Table 1. In this population, there are two clusters of size 3. The 1st, 2nd, and 3rd units form cluster 1 consisting of a network (1st and 2nd) and an edge unit (3rd unit). The 5th, 6th, and 7th form cluster 2 consisting of a network (6th and 7th) and an edge unit (5th unit). For

example, in the fourth row of Table 1, the 1st, 2nd, and 6th units were selected as the initial sample. Since all of them exceed 10, their neighboring units were also added to the sample. Thus,

$$d_J = \{(1, 12, 1), (2, 1000, 1), (3, 4, 0), (5, 5, 0), (6, 500, 1), (7, 30, 1)\}.$$

Ignoring the edge units, the computations for the HH and HT estimators are $w_1 = w_2 = (12 + 1000)/2 = 506$ and $w_3 = (500 + 30)/2 = 265$, leading to

$$\hat{\mu}_{HH} = \frac{(506 + 506 + 265)}{3} = 425.67,$$

and for $\hat{\mu}_{HT}$ we need, $y_1^* = 1012$, $y_2^* = 530$, $\alpha_1 = \alpha_2 = 1 - \binom{5}{3}/\binom{7}{3} = 0.71$, leading to

$$\hat{\mu}_{HT} = \left[\frac{\left(\frac{1012}{0.71}\right) + \left(\frac{530}{0.71}\right)}{7} \right] = 308.40.$$

The computations for $\tilde{\mu}_{HH}$ are $m. = 2 + 2$, $\xi = \binom{4}{3} - 0 - 0 = 4$, $\xi_1 = \xi_2 = \binom{3}{2} - 0 - 0 = 3$. From (11),

$$\tilde{\mu}_{HH} = \frac{(12 \times 3 + 1000 \times 3 + 500 \times 3 + 30 \times 3)}{(3 \times 4)} = 385.5.$$

From (8) $\widehat{\text{var}}(\hat{\mu}_{HH}) = 3687.68$, and $\xi_{12} = \xi_{13} = \xi_{23} = \binom{2}{1} - 0 - 0 = 2$. Then, from (12),

$$\begin{aligned} \text{var}[\hat{\mu}_{HH}|d_J] &= \frac{1}{3^2 \times 4^2} [(2 \times 506^2 + 2 \times 265^2)(4 \times 3 - 3^2) \\ &\quad + 2(506 \times 506 + 4 \times 506 \times 265 + 265 \times 265)(2 \times 4 - 3^2)] \\ &= 1613.36. \end{aligned}$$

Hence, $\widehat{\text{var}}(\tilde{\mu}_{HH}) = 3687.68 - 1613.36 = 2074.32$.

Example 2 (Point-objects) This example is introduced to demonstrate how we take advantage of the situation when most of the selected networks are of size one to reduce the computations of ξ , ξ_i and ξ_{ij} . This example is from Thompson (1990). The population of point-objects illustrated in Fig. 2 was produced as a realization in the unit square of a Poisson cluster process (cf. Diggle, 1983) with five ‘‘parent locations’’ from a uniform distribution, and a random (Poisson mean = 40) number of point-objects dispersed in relation to the parent locations with a symmetric Gaussian distribution having standard deviation 0.02. The population mean is $\mu = 190/400 = 0.475$. The neighborhood of a unit consists of the unit itself and the four adjacent units sharing a common boundary line.

In Fig. 2, the initial sample consists of 15 units selected by simple random sampling from 400 units. Two of the initial sample units, near the top of the study region, landed in a network of size $m_1 = 6$ with $w_1 = 6$. Two other initial samples landed in a network of size $m_2 = 11$ with $w_2 = 9.27$. The other eleven units landed in networks of size one with $w_i = 0$. Thus, $m^* = 11 + 6 = 17$, $n_1^* = 2 + 2 = 4$, and $\kappa^* = 2$. Now using the formula for ξ ,

$$\xi = \binom{17}{4} - \binom{11}{4} - \binom{6}{4} = 2035.$$

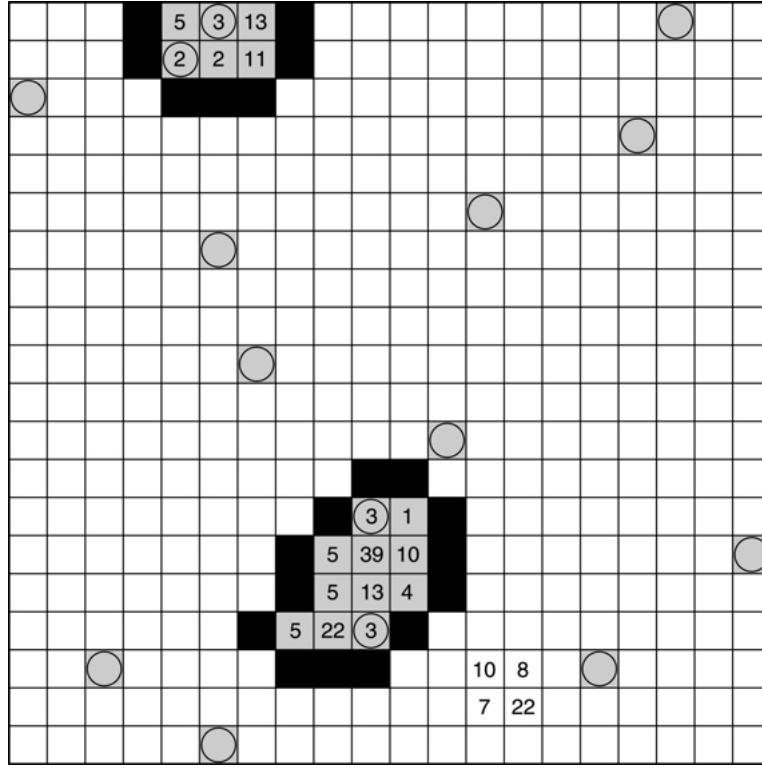


Figure 2. The numbers in each unit denote the number of objects. The units without a number denote units without objects. An initial simple random sample of 15 units is selected, which are shown as circles. Adaptively adding the units, the 13 selected networks are shaded regions. The black regions are selected edge units.

For the units in the first network

$$\xi_1 = \binom{16}{3} - \binom{5}{3} = 550 \quad \text{and} \quad \xi_2 = \binom{16}{3} - \binom{10}{3} = 440.$$

Using (10), $\hat{\mu}_{HH} = 2.19$. From (8), $\widehat{\text{var}}(\tilde{\mu}_{HH}) = 0.895$, also $\xi_{ij} = \binom{15}{2} - \binom{4}{2} = 99$ when i and j are in network 1, $\xi_{ij} = \binom{15}{2} - \binom{9}{2} = 69$ when i and j are in network 2, and $\xi_{ij} = \binom{15}{2} = 105$ when i is in network 1 and j is in network 2. From (12) $\widehat{\text{var}}[\tilde{\mu}_{HH}] = 0.895 - 0.028 = 0.867$.

For $\binom{6}{1}\binom{11}{3}$, $\binom{6}{3}\binom{11}{1}$ and $\binom{6}{2}\binom{11}{2}$ numbers of the possible initial samples giving rise to d_j , we respectively have $\widehat{\text{var}}[\tilde{\mu}_{HH}] = 1.088, 0.694$ and 0.895 . From (12),

$$\widehat{\text{var}}_{RB}[\hat{\mu}_{HH}] = \frac{1}{2035} \left[\binom{6}{1}\binom{11}{3} 1.088 + \binom{6}{3}\binom{11}{1} 0.694 + \binom{6}{2}\binom{11}{2} 0.895 \right] = 0.967.$$

Hence, $\widehat{\text{var}}_{RB}[\tilde{\mu}_{HH}] = 0.967 - 0.028 = 0.939$.

In order to shed some light on the properties of $\tilde{\mu}_{HH}$, we simulated this example. For $n_1 = 10, 20, 30, 40, 50, 60, 100, 200$, we ran a simulation of 10,000 samples. For each

Table 2. Variance of $\hat{\mu}_{HT}$, $\tilde{\mu}_{HH}$, and $\hat{\mu}_{HT}$ with initial sample size n_1 for the population illustrated in Fig. 2.

n_1	$var[\hat{\mu}_{HT}]$	$var[\tilde{\mu}_{HH}]$	$var[\hat{\mu}_{HT}]$
10	0.38655	0.41899	0.42001
20	0.17097	0.19996	0.20462
30	0.10030	0.13167	0.13282
40	0.06587	0.09655	0.09693
50	0.04593	0.07334	0.07539
60	0.03322	0.06028	0.06103
100	0.01096	0.03174	0.03231
200	0.00106	0.01040	0.01077

replication, $\tilde{\mu}_{HH,i}$ was calculated where the index i stands for replication number. We calculate

$$MSE(\tilde{\mu}_{HH}) = \frac{1}{r-1} \sum_{i=1}^r (\tilde{\mu}_{HH,i} - \bar{\mu}_{HH})^2 + (\bar{\mu}_{HH} - \mu)^2,$$

where $r(= 10,000)$ is the number of replications and $\bar{\mu}_{HH} = \frac{1}{r} \sum_{i=1}^r \tilde{\mu}_{HH,i}$. The values of $(\bar{\mu}_{HH} - \mu)^2$ were negligible for all n_1 . However, we consider $MSE(\tilde{\mu}_{HH})$ as the variance of $\tilde{\mu}_{HH}$. The results are given in Table 2. We see that $\hat{\mu}_{HT}$ is more efficient than $\tilde{\mu}_{HH}$ in this example.

5. Discussion

The information from the edge units are not used in $\hat{\mu}_{HH}$ and $\hat{\mu}_{HT}$. Thompson and Seber (1996) noted that the decision to add adaptively can be based on a cheap variable, or so-called ‘‘rapid assessment variable’’. The information from the rapid variable cannot improve the estimators using Rao-Blackwell theory. Therefore, in this case, $\hat{\mu}_{HT}$ is an admissible estimator (the same as Horvitz-Thompson estimator for conventional sampling designs) in the class of unbiased estimators (Godambe and Joshi, 1965) whereas $\hat{\mu}_{HH}$ remains an inadmissible estimator. In Section 3, we showed that $\hat{\mu}_{HT}$ has the advantage of holding the appropriate proportionality property more closely than $\hat{\mu}_{HH}$ for clumped and rare populations. In Section 4, we also showed that $\hat{\mu}_{HT}$ has the advantage of not being dominated, in the sense of having smaller variance, by any estimator, but $\hat{\mu}_{HH}$ is dominated by $\tilde{\mu}_{HH}$. For two examples in Section 4, $\hat{\mu}_{HT}$ is even more efficient than $\tilde{\mu}_{HH}$. Therefore, ecologists are strongly encouraged to use $\hat{\mu}_{HT}$ rather than $\hat{\mu}_{HH}$. However, a comprehensive study is needed to compare $\hat{\mu}_{HT}$ and $\tilde{\mu}_{HH}$.

Acknowledgment

This research is supported by the Research Council of Isfahan University of Technology. The author is very grateful to Dr David Smith for valuable comments and suggestions that

enhanced the quality of this paper. He also thanks the anonymous referees for their suggestions which improved the presentation of the paper.

References

- Basu, D. (1969) Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā A*, **31**, 441–54.
- Brown, J.A. (1994) The application of adaptive cluster sampling to ecological studies. In D.J. Fletcher and B.F.J. Manly (eds), *Statistics in Ecology and Environmental Monitoring*, **2**, 86–97, Dunedin, University of Otago Press, New Zealand.
- Brown, J.A. and Manly, B. (1998) Restricted adaptive cluster sampling. *Journal of Environmental and Ecological Statistics*, **5**, 49–63.
- Diggle, P.J. (1983) *Statistical Analysis of Spatial Point Patterns*, Academic Press, New York.
- Godambe, V.P. and Joshi, V.M. (1965) Admissibility and Bayes estimation in sampling finite population I. *Annals of Mathematical Statistics*, **36**, 1701–22.
- Lo, N., Griffith, D., and Hunter, J.R. (1997) Using a restricted adaptive cluster sampling to estimate Pacific hake larval abundance. *CalCOFI Rep.* **38**, 103–13.
- Quinn II, T.J., Hanselman, D.H., Clausen, D.M., Heifetz, J., and Lunsford, C. (1999) Adaptive cluster sampling of rockfish populations. *Proceedings of the American Statistical Association*, Joint Statistical Meetings, Biometrics Section, 11–20.
- Roesch, Jr., F.A. (1993) Adaptive cluster sampling for forest inventories. *Forest Science*, **39**, 655–69.
- Salehi, M.M. and Seber, G.A.F. (1997a) Adaptive cluster sampling with networks selected without replacement. *Biometrika*, **84**, 209–19.
- Salehi, M.M. and Seber, G.A.F. (1997b) Two-stage adaptive cluster sampling. *Biometrics*, **53**, 959–70.
- Salehi, M.M. and Seber, G.A.F. (2002) Unbiased estimators for restricted adaptive cluster sampling with networks selected with and without replacement. *Australian and New Zealand Journal of Statistics*, **44**, 63–74.
- Särndal, C.E., Swensson, B., and Wretman, J. (1992) *Model Assisted Survey Sampling*, Springer-Verlag, New York.
- Smith, D.R., Conroy, M.J., and Brakhage, D.H. (1995) Efficiency of adaptive cluster sampling for estimating density of wintering waterfowl. *Biometrics*, **51**, 777–88.
- Thompson, S.K. (1990) Adaptive cluster sampling. *Journal of the American Statistical Association*, **85**, 1050–9.
- Thompson, S.K. (1991a) Adaptive cluster sampling: Designs with primary and secondary units. *Biometrics*, **47**, 1103–15.
- Thompson, S.K. (1991b) Stratified adaptive cluster sampling. *Biometrika*, **78**, 389–97.
- Thompson, S.K. (1996) Adaptive cluster sampling based on order statistics. *Environmetrics*, **7**, 123–33.
- Thompson, S.K. and Seber, G.A.F. (1996) *Adaptive Sampling*, Wiley New York.

Biographical sketch

Mohammad Salehi Marzijarani was a lecturer in Department of Mathematics, Arak University, Arak, Iran, from 1990 to 1994. He has got his Ph.D. in Statistics at Auckland University, New Zealand in 1998. He is an assistant professor in the School of Mathematical Sciences, Isfahan University of Technology, Isfahan, Iran.

Copyright of Environmental & Ecological Statistics is the property of Kluwer Academic Publishing / Academic and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.