HORVITZ–THOMPSON THEOREM AS A TOOL FOR GENERALISATION OF PROBABILITY SAMPLING TECHNIQUES

I.A Adeleke

Department of Actuarial Science and Insurance,
University of Lagos, Lagos Nigeria
Email: adeleke22000@yahoo.ca

E.O. Esan

Department of Mathematics,
University of Lagos, Lagos, Nigeria
Email: esaneo@yahoo.com

Ray O. Okafor

Department of Mathematics,
University of Lagos, Lagos Nigeria
Email: okaforray@yahoo.com

ABSTRACT

The Horvitz –Thompson estimator is often presented in the context of an estimator based on the concept of variable probability sampling. In truth, however, the Horvitz –Thompson theorem is arguably a lot more versatile than this narrow perspective. The main objective of this article is to review the application of the generalization provided by the Horvitz –Thompson theorem to basic sampling methodology, and to estimators of means and variances. It is further shown that the Horvitz –Thompson estimator also serves as the basis for a general strategy for consistent estimation.

KEY DESCRIPTORS: Horvitz –Thompson Theorem, Variable Probability Sampling, Generalization of Probability Sampling, Consistent Estimation

INTRODUCTION

An earlier observation by Stuart (1962) that "Sample Survey Theory seems, more than most branches of statistics, to suffer the lack of a unifying thread of statistics, on which to string the various sampling techniques of which it is composed," was shared by many students of sample survey after completing the course. The Horvitz – Thompson Theorem, which most sampling texts often introduced in the narrow context of variable probability sampling of primary sampling units (cluster) according to a measure of size of the cluster (Cochran, 1977) offers a needed integrating perspective for **teaching** sampling methods and estimating population total and population variance. The paper emphasizes the importance of the Horvitz–Thompson Theorem and its role as a unifying theory of probability sampling by focusing on Inclusion Probabilities and the Hovitz-Thompson Theorem.

INCLUSION PROBABILITIES

Inclusion probabilities provide a natural transition from the sample space representation to the inclusion probability representation used in Horvitz -Thompson. The first-order inclusion probability, $\pi_u = \sum_{u \in S} P(S)$, defined for each element u of the finite universe, U is the probability that element u will be included in the sample. The second-order, or pair wise inclusion probability, denoted by π_{uv}, is the probability that units' u and v will be included in the sample, and is calculated by $\pi_{uv} = \sum_{S \ni u,v} P(S)$ with summation now overall samples containing both elements u and v.

These inclusion probabilities are determined by the designs and can be specified without reference to a response variable. For many designs the required inclusion probabilities are readily calculated without the function $P(S)$.

The inclusion probabilities for three common designs are stated below:

1). Simple random sampling (SRS):

$$\pi_u = \frac{n}{N} \tag{1}$$

$$\pi_{uv} = \frac{n(n-1)}{N(N-1)} \quad \text{\{Selection without replacement\}} \tag{2}$$

2). Stratified random sampling; selecting a SRS of n_h element from the N_h elements in stratum h:

$$\pi_u = \frac{n_h}{N_h} \quad \text{for u in stratum } h \tag{3}$$

$$\pi_{uv} = \frac{n_h(n_h-1)}{N_h(N_h-1)} \quad \text{if } u \text{ and } v \text{ are both in stratum } h$$

$$= \pi_u \pi_v \quad \text{if } u \text{ and } v \text{ are in different strata} \tag{4}$$

3). Systematic sampling with a random start on (1,…,k) and sampling interval k, where k is an integer:

$$\pi_u = \frac{1}{k} \quad \text{\{Equal probability out of } k\text{\}} \tag{5}$$

$$\pi_{uv} = \frac{1}{k} \quad \text{if } u \text{ and } v \text{ occur in the same sample}$$

$$= 0 \quad \text{if } u \text{ and } v \text{ do not occur in the same sample} \tag{6}$$

Once one has gained familiarity with the basic sampling designs and estimators of means, totals, and proportions in the sample space presentation, one may now be introduced to the unifying perspective provided by the Horvitz–Thompson and the inclusion probability representation of the sampling designs (Overton and Stehman, 1995).

THE HORVITZ-THOMPSON THEOREM

Defining the Theorem

For strictly positive probabilities i.e. $\pi_i > 0$ for every $i \in N$, the Horvitz – Thompson estimator

$$\hat{Y}_{HT} = \frac{\sum_{i=1}^{N} y_i}{\pi_i} \tag{7}$$

is unbiased for the population total $Y_{HT} = \sum_{i=1}^{N} y_i$ with variance

$$\operatorname{var}(\hat{Y}_{HT}) = \sum_{i=1}^{N}\left(\frac{y_i}{\pi_i}\right)^2 (1-\pi_i) + 2\sum_{i=1}^{N}\sum_{j=i+1}^{N}\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j} y_i y_j \tag{8}$$

The Horvitz–Thompson theorem may be used to derive estimators of means and variances for many commonly used designs (Horvitz & Thompson, 1952; Overton & Stehman 1992).

The Horvitz–Thompson theorem, which is a design-based approach to survey inference, has a number of strengths that make it popular with practitioners. It automatically takes into account features of the survey design, and provides reliable inferences in large samples, without the need for strong modelling assumptions. On the other hand, it is essentially asymptotic, and hence yields limited guidance for small-sample adjustments. It lacks a theory for optimal estimation (Godambe & Thompson, 1986) and estimates from the approach are potentially inefficient.

Consider inference about the population total

$$T = Y_1 + \ldots + Y_N$$

and any sample design with positive inclusion probability

$\pi_i = E(I_i \mid y) > 0$ for units i, $i = 1, 2, \ldots, N$. The HT estimator is then

$$\hat{T}_{HT} = \sum_{i=1}^{n} y_i / \pi_i = \sum_{i=1}^{N} I_i y_i / \pi_i \qquad (9)$$

and is design unbiased for T, since

$$E(\hat{T}_{HT} \mid y) = \sum_{i=1}^{N} E(I_i \mid y_i) / \pi_i = \sum_{i=1}^{N} y_i$$

The unbiasedness of (9) under very mild conditions conveys robustness to modelling assumptions, and makes it a mainstay of the design-based approach. But (9) has two major deficiencies. First, the choice of variance estimator is problematic for some designs. Second, the HT estimator can have a high variance. For example, when an outlier in the sample has a low selection probability and hence receives a large weight, Basu's (1971) famous circles elephant example provides an amusing, if extreme example.

Due to the emphasis on equal probability sampling to which many people are normally exposed in introductory statistics, they may initially view unequal probability sampling as unacceptable or even "biased". However, the Horvitz – Thompson theorem establishes the intuitively appealing solution that unequal inclusion probabilities are accounted for simply by using the appropriate weights, the inverse of the inclusion probabilities, in the estimators. Similarly, Stuart's (1962) used the phrase "apparent frequency", in reference to the weight $w_u = 1/\pi_u$, indicates that an element of the sample with inclusion probability π_u represents w_u elements of U when the sample data are "expanded" to estimate totals and means over U.

Consistency of the Horvitz-Thompson Theorem

Overton and Stehman (1995) discussed the consistency of the Horvitz–Thompson theorem by applying an adaptation of Fisher's (1956) definition of consistency. This consistency result establishes the Horvitz – Thompson estimator as the basis for a generalized strategy of consistent estimation: any function of the linear parameters is estimated consistently by the same function of Horvitz–Thompson estimators of those linear parameters. This result is an important part of sampling methodology. For a sampling methods course, it may be sufficient to mention consistency as an important property for an estimator to posses, and then to present several applications of this method for constructing estimators.

Examples

A. <u>Simple Random Sampling</u>:

Consider taking a sample of size four from a population of size 10 as indicated below, where y_i's are the observed units i.e.

$$N = 10 \; and \; n = 4$$

$$with \; y_i = 8, 2, 4, 3$$

$$\pi_i = \frac{n}{N} = \frac{4}{10} = 0.4$$

$$\pi_{ij} = 0.16 \; \{ \; there \;\; is \;\; equal \;\; probability \; of \;\; selection \;\; for \;\; all \;\; units \},$$

$$\pi_i^{\,2} = 0.16, \; \pi_i \pi_j = \frac{4}{10} \times \frac{3}{9} = \frac{4}{30} \; \{as \;\; sampling \;\; is \;\; done \;\; without \;\; replacement\}$$

with Horvitz–Thompson Theorem, we obtained

1. Estimate of the population total

$$\hat{Y}_{HT,srs} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \frac{17}{0.4} = 42.5$$

Estimate of the population variance

$$\mathrm{var}(\hat{Y}_{HT,srs}) = \sum_{i=1}^{4}\left[\frac{1}{\pi_i^2}-\frac{1}{\pi_i}\right]y_i^2 + 2\sum_{i=1}^{4}\sum_{j=2}^{4}\left[\frac{1}{\pi_i\pi_j}-\frac{1}{\pi_{ij}}\right]y_iy_j$$

$$= \quad S_1 \quad + \quad 2S_2$$

$$S_1 = 348.75,\ S_2 = -122.5$$

$$\therefore \mathrm{var}(\hat{Y}_{HT,srs}) = S_1 + 2S_2$$

$$= 103.75$$

with the common method, estimate of the population total is given by

Estimate of the population variance is given as

$$\mathrm{var}(\hat{Y}_{srs}) = N^2\,\mathrm{var}(\hat{\bar{y}})$$

$$= N^2\left(\frac{1-f}{n}\right)s^2$$

$$\therefore \mathrm{var}(\hat{Y}_{srs}) = 103.75\ \{\mathit{same\ as\ HT's}\}$$

B. Stratified Sampling

Consider the stratification of a population of size N = 57 into three strata which population sizes N_1, N_2, N_3 respectively as given below with their respective observed units from each stratum;

$N_1=12$, $N_2=20$, $N_3=25$
$y_{i1} = 4, 5, 6$ {units in stratum 1}
$y_{i2} = 7, 6, 4, 3$ {units in stratum 2}
$y_{i3} = 5, 3, 2, 1$ {units in stratum 3}

Using the Horvitz–Thompson Theorem we obtained the

1. Estimate of the population total thus:

$$\hat{Y}_{HT,st} = \sum_{h=1}^{l}\sum_{i=1}^{n_h}\frac{y_{hi}}{\pi_{hi}}$$

$$\pi_{hi} = \frac{n_h}{N_h}$$

$$\therefore \hat{Y}_{HT,st} = \frac{15}{0.25}+\frac{20}{0.2}+\frac{25}{0.33} = 235$$

2. Estimate of the population variance is obtained as:

$$\text{var}(\hat{Y}_{HT,st}) = \sum_{h=1}^{l}\sum_{i=1}^{nh}\left(\frac{1}{\pi_{hi}^{2}} - \frac{1}{\pi_{hi}}\right)y_{hi}^{2} + 2\sum_{h=1}^{l}\sum_{i=1}^{nh}\sum_{j>i}^{nh}\left(\frac{1}{\pi_{hi}\pi_{hj}} - \frac{1}{\pi_{hij}}\right)y_{hi}y_{hj}$$

when $h = 1$ (*stratum* 1)

$$\pi_i = \frac{1}{4} = 0.25, \quad \pi_i^{2} = 0.0625, \quad \pi_i\pi_j = 0.0625, \quad \pi_{ij} = \frac{3}{12}\times\frac{2}{11} = \frac{1}{22}$$

$y_i = 4, 5, 6$

$$\text{var}(\hat{Y}_{THT,st}) = \sum\left(\frac{1}{\pi_i^{2}} - \frac{1}{\pi i}\right)y_i^{2} + 2\sum\sum\left(\frac{1}{\pi_i\pi_j} - \frac{1}{\pi_{ij}}\right)y_iy_j$$

$$= S_1 \qquad + \qquad 2S_2$$

$S_1 = 924$

$2S_2 = -888$

$\therefore \text{var}(\hat{Y}_{HT,st1}) = 36$

when $h = 2$ (*stratum* 2)

$$\pi_i = 0.2, \quad \pi_i^{2} = 0.04, \quad \pi_i\pi_j = 0.04, \quad \pi_{ij} = \frac{12}{380}$$

$y_i = 7, 6, 4, 3$

$$\text{var}(\hat{Y}_{THT,st}) = \sum\left(\frac{1}{\pi_i^{2}} - \frac{1}{\pi i}\right)y_i^{2} + 2\sum\sum\left(\frac{1}{\pi_i\pi_j} - \frac{1}{\pi_{ij}}\right)y_iy_j$$

$$= S_1 \qquad + \qquad 2S_2$$

$S_1 = (25 - 5)(7^2 + 6^2 + 4^2 + 3^2)$

$\quad = 2200$

$S_2 = (25 - \frac{380}{12})(7\times 6 + 7\times 4 + 7\times 3 + 6\times 4 + 6\times 3 + 4\times 3) = -\frac{11600}{12}$

$2S_2 = -1933.33$

$\therefore \operatorname{var}(\hat{Y}_{HT,st2}) = 2200 - 1933.33 = 266.67$

when $h = 3$ (*stratum* 3)

$$\pi_i = \frac{1}{5} = 0.2, \quad \pi_i^2 = 0.04, \quad \pi_i \pi_j = 0.04, \quad \pi_{ij} = \frac{5}{25} \times \frac{4}{24} = \frac{1}{30}$$

$y_i = 5, 3, 2, 1, ,4$

$$\operatorname{var}(\hat{Y}_{THT,st}) = \sum \left(\frac{1}{\pi_i^2} - \frac{1}{\pi i} \right) y_i^2 + 2 \sum \sum \left(\frac{1}{\pi_i \pi_j} - \frac{1}{\pi_{ij}} \right) y_i y_j$$

$$= S_1 \qquad + \qquad 2S_2$$

$S_1 = 1100$

$2S_2 = -850$

$\therefore \operatorname{var}(\hat{Y}_{HT,st3}) = 250$

$\Rightarrow \operatorname{var}(\hat{Y}_{HT,st}) = 552.67$ {var *iance for the entire population* }

with the common method we have

$N_1 = 12$ $\qquad$ $N_2 = 20$ $\qquad$ $N_3 = 25$

Stratum I		Stratum II		Stratum III	
y	y^2	Y	y^2	y	y^2
4	16	7	49	5	25
5	25	6	36	3	9
6	36	4	16	2	4
15	77	3	9	1	1
$\bar{y} = 5$		20	110	4	16
		$\bar{x} = 5$		15	55
				$\bar{x} = 3$	

Estimate of the population total

$$\hat{Y}_{st} = N \hat{\bar{y}}_{st}$$

where ,

$$\hat{\bar{y}}_{st} = \sum W_h \bar{y}_h = \frac{235}{57}$$

$$\therefore \hat{Y}_{st} = 57 \times \frac{235}{57} = 235$$

Estimate of the population variance

$$\text{var}(\hat{Y}_{st}) = N^2 \,\text{var}(\hat{\bar{y}}_{st})$$

$$= N^2 \sum_h \frac{(1-f_h)s_h^{\,2}}{n_h} W_h$$

for stratum 1, we have

$$s_1^{\,2} = 1, \quad f_1 = \frac{n_1}{N_1} = \frac{1}{4}, \quad 1 - f_1 = \frac{3}{4}, \quad n_1 = 3$$

for stratum 2, we have

$$s_2^{\,2} = \frac{10}{3}, \quad f_2 = \frac{n_2}{N_2} = \frac{1}{5}, \quad 1 - f_2 = \frac{4}{5}, \quad n_2 = 4$$

for stratum 3, we have

$$s_3^{\,2} = \frac{10}{4}, \quad f_3 = \frac{n_3}{N_3} = \frac{1}{5}, \quad 1 - f_3 = \frac{4}{5}, \quad n_3 = 5$$

$$\Rightarrow \text{var}(\hat{Y}_{st}) = 552.67$$

C. Adaptive Cluster Sampling

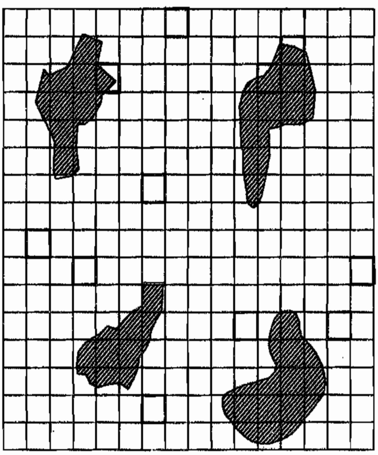
Adaptive sampling designs are designs in which additional units or sites for observation are selected depending on the interpretation of observations made during the survey. Additional sampling is driven by the observed results from an initial sample. In the simplest form of adaptive cluster sampling an initial sample of units is selected by random sampling with or without replacement (Thompson 1990) and whenever the variable of interest for a unit in the sample satisfies a pre-specified condition, neighboring or connected units are added to the sample and observed (Thompson 1997). The usual design unbiased estimators for adaptive cluster sampling with initial sample taken by with or without replacement are of a Horvitz-Thompson type.

Figure 1 (showing how adaptive cluster samples are drawn from a population) below, illustrates a network and its associated edge units, which together will be called a cluster. Shaded areas on the figure indicate the area of interest; for instance, areas of elevated contaminant levels. This example has four regions of contamination. The 12 darkened rectangles in the figure represent a randomly selected set of 12 sampling units constituting the initial sample. Whenever a sampled unit is found to exhibit the characteristic of interest — that is, the unit intersects any part of the shaded areas —

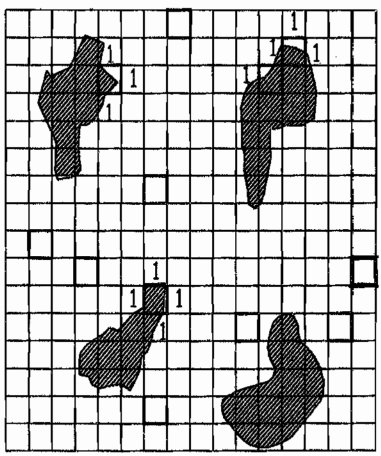
neighboring sampling units are also sampled using a consistent pattern. An example of follow-up sampling pattern is shown in Figure 2, where the x's indicate the neighboring sampling units to be sampled. The follow-up sampling pattern is called the *neighborhood* of a sampling unit. The five grid units in the figure make up the neighborhood of the initially sampled unit. In Figure 1(a), four initial sampling units intersect the shaded areas. The units adjacent to these four initial units are sampled next, as shown in Figure 1(b). Some of these sampled adjacent units also intersect the shaded areas, so the units adjacent to these are sampled next, as shown in Figure 1(c). Figures 1(d) to (f) show subsequent sampling until no more sampled units intersect the shaded areas. Figure 3 shows the initial random sample and the final sample. Note that the final sample covers three of the four regions of contamination. If at least one of the initial units had intersected the fourth area, it would also have been covered by a cluster of observed units.

The final sample consists of *clusters* of selected (observed) units around the initial observed units. Each cluster is bounded by a set of observed units that do not exhibit the characteristic of interest. These are called *edge units*. A cluster without its edge units is called a *network*. Any observed unit, including an edge unit, that does not exhibit the characteristic of interest is a network of size one. Hence, the final sample can be partitioned into non-overlapping networks. These definitions are important in understanding the estimators for statistical parameters.

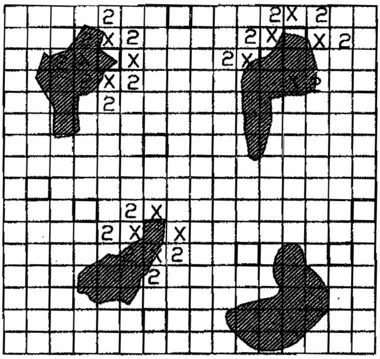
Fig. 1: Population Grid with Shaded Areas of Interest, Initial Simple Random Sample, and Follow-up Sample



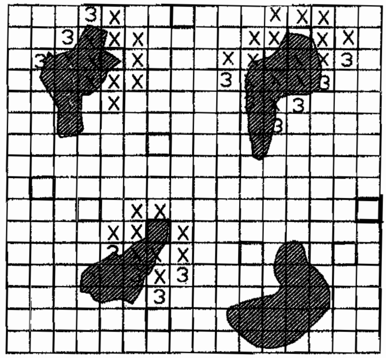
(a) Initial sample



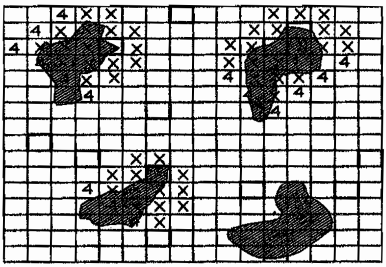
(b) First batch of adjacent units



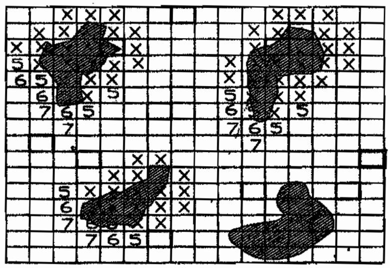
(c) Second batch of adjacent units



(d) Third batch of adjacent units



(e) Fourth batch of adjacent units



(f) Fifth, sixth and seventh batch of adjacent units

Fig. 2: Follow-up Sampling Pattern

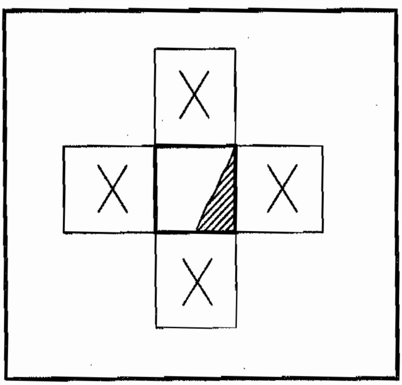
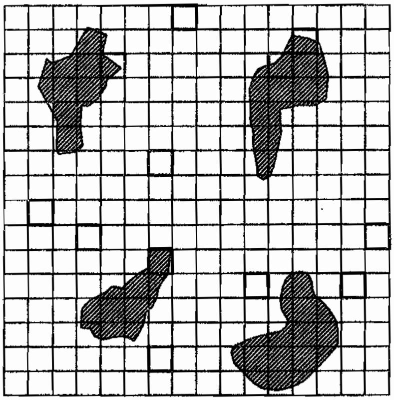
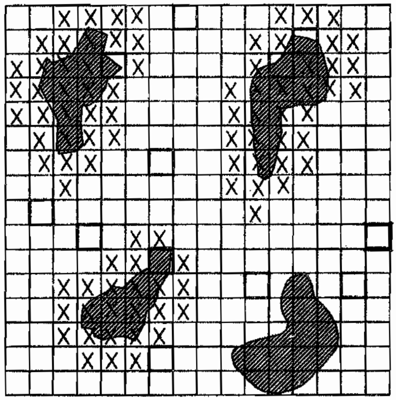


Fig. 3: Population Grid with Shaded Areas of Interest, Initial Simple Random Sample, and Final Sample



Population Grid with Shaded Areas of Interest and Initial Simple Random Sample



Final Adaptive Cluster Sampling Results

X = Observed Sampling unit

Estimators of mean and variance for adaptive cluster sampling

For an adaptive cluster sample with an initial simple random sample, the modified Horvitz-Thompson forms of the estimators are

$$\hat{\mu} = \frac{1}{N}\sum_{k=1}^{K}\frac{y_k^*}{\alpha_k}$$

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

where y_k^* = sum of the values of the character of interest, y, for the k^{th} network in the sample

N = number of units in the population

K= number of distinct networks in the sample

α_k = probability that the initial sample intersects the k^{th} network

α_{jk} = probability that the initial sample intersects both the j^{th} and the k^{th} networks

Units in the initial sample that do not satisfy the pre-specified condition (i.e. not intersecting the shaded area) are included in the calculation as networks of size one, but edge units are excluded.

If there are x_k units in the k^{th} network, then the intersection probabilities α_k and α_{jk} are calculated using combinatorial formulas as follows:

$$\alpha_k = 1 - \left[\binom{N - x_k}{n_1} \Big/ \binom{N}{n_1}\right]$$

$$\alpha_{jk} = 1 - \left[\left[\binom{N - x_j}{n_1} + \binom{N - x_k}{n_1} - \binom{N - x_j - x_k}{n_1}\right] \Big/ \binom{N}{n_1}\right]$$

where $a_{jj} = a_j$

Application (Using the Horvitz – Thompson Estimators)

Consider the adaptive cluster sample shown in Figure 1. There are N=256 grid units in the population and $n1$=12 units in the initial sample. One initial sample unit on the upper left area of the study region intersected a network of $x1$=18 units. Let this be network A1. Two other initial sample units on the upper right area of the study region intersected a network (A2) of $x2$=19 units. Another initial sample intersected the network of x_3=13. Let this network be A_3. The remaining seven initial sample units form networks of size one (A_4,...,A_{11}). Hence, there are k=11 distinct networks, with $x1$=18, $x2$=19, $x3$=13,
x_4 = $x5$ =...= $x11$ =1 units, respectively. The intersection probabilities are calculated to be
α_1= 0.591396, α_2= 0.611997, α_3 = 0.472509 and for the remaining networks (A4, A5,...,A11) the intersection probability is α_k = 0.046875 for k = 4, 5, to 11
Hence, the estimate of the mean using the Horvitz-Thompson estimator is

where y_1^* is the sum of the 18 observations from network A1, y_2^* is the sum of the 19 observations from network A_2, y_3^* is the sum of the 13 observations from the network A_3, and y_4^*, y_5^*, ..., y_{11}^* are the single observations from the networks of size one.

To compute an estimate of the variance, we need the joint intersection probabilities which can easily be shown to be:

$\alpha_{12} = 0.35025$, $\alpha_{13} = 0.268724$, $\alpha_{23} = 0.278403$, $\alpha_{1j} = 0.026273$ for $j = 4, \ldots, 11$

$\alpha_{2j} = 0.027229$ for $j = 4, \ldots, 11$, $\alpha_{3j} = 0.020826$ for $j = 4, \ldots, 11$,

$\alpha_{jk} = 0.002022$ for $j = 4, \ldots, 11$ and $k = 4, \ldots, 11$, $j \neq k$

Then the variance using the HT is ,

$$= \frac{1}{(256)^2}\left[\frac{(y_1^*)^2}{\alpha_1}\left(\frac{1}{\alpha_1} - 1\right) + \ldots + \frac{(y_{11}^*)^2}{\alpha_{11}}\left(\frac{1}{\alpha_{11}} - 1\right)\right]$$

$$+ \frac{1}{(256)^2}\left[2\frac{(y_1^* y_2^*)}{\alpha_{12}}\left(\frac{\alpha_{12}}{\alpha_1\alpha_2} - 1\right) + \ldots + 2\frac{(y_{10}^* y_{11}^*)}{\alpha_{10,11}}\left(\frac{\alpha_{10,11}}{\alpha_{10}\alpha_{11}} - 1\right)\right]$$

CONCLUSION

The Horvitz – Thompson theorem offers a needed integrating perspective for understanding the methods and fundamental concepts of probability sampling in sample survey. Development of basic concepts in sampling via this approach would provide researchers and other users with the tools to solve more complicated problems, and helps to avoid some common stumbling blocks of estimation of population parameters when units have varying probability of selection. Our submission is that a greater appreciation of the Horvitz–Thompson will contribute a beneficial integrated perspective that assists in presenting sampling.

REFERENCES

Basu, D. (1971). "An Essay on the Logical Foundations of Survey Sampling, Part 1." Foundations of Statistical Inference, Toronto: Holt, Richard and Winston. 203-242.

Cochran, W.G. (1977). Sampling Techniques (3rd ed.), New York, NY: John Wiley.

Fisher, R. A. (1956). Statistical Methods and Scientific Inference (3rd ed.), New York, NY: Hafner Press.

Godambe, V. P. & Thompson, M. E. (1986). "Parameters of Superpopulations and Survey Population: their Relationship and Estimation." International Statistical Review 54: 37-59

Horvitz, D.G. and Thompson, D.J. (1952). "A Generalization of Sampling without Replacement from a Finite Universe." Journal of the American Statistical Association, 47: 663-685

Thompson, S.K. (1990). "Adaptive Cluster Sampling." Journal of the American Statistical Association 85: 1050-1059.

Thompson, S.K. (1992). Sampling. New York: Wiley.

Thompson, S.K. & Seber, G.A.F. (1996). Adaptive Sampling. New York: Wiley.

Thompson, S. K. (1997). "Adaptive Sampling in Behavioral Surveys." In Harrison, L., and Hughes, A. (eds.). The Validity of Self-Reported Drug Use: Improving the Accuracy of Survey Estimates.(NIDA Research Monograph 167). Rockville, MD

Overton, W.S. & Stehman, S. V. (1992). "The Horvitz – Thompson Theorem as a Unifying Perspective for Sampling," in Proceedings of the Section on Statistics Education of the American Statistical Association, Alexandria, VA: American Statistical Association: 380-385.

Overton, W. S. & Stehman, S. V. (1995). "The Horvitz – Thompson Theorem as a Unifying Perspective for Sampling: with Examples form Natural Resource Sampling" American Statistical Association.

Stuart, A. (1962). The Basic Ideas of Scientific Sampling (1984 ed.), London: Charles Griffin and Company.