J. Ind. Soc. Agril. Statist.
58(2), 2004 : 177-189

Five Decades of the Horvitz-Thompson Estimator and Furthermore...

T.J. Rao
Indian Statistical Institute, Kolkata

SUMMARY

We shall first trace the historical aspects of the Horvitz-Thompson estimator and quickly review some of the important optimality properties. We shall then discuss a few new areas of application where HT estimator plays an important role.

Key words : Harvitz-Thompson estimator, Optimality of HT estimator.

1. Introduction

While drawing inferences based on samples from a finite population of labelled units, one comes across a natural deviation from the classical theory of estimation for the case of infinite populations, wherein a linear estimator for a sample of n units is defined as $\Sigma_{i=1}^{n} \omega_i y_i$, ω_i's are independent of observations y_i's and $\bar{y} = \frac{1}{n} \Sigma_{i=1}^{n} y_i$ is the m.v.u.e. of $E(y_i)$ in the class of all linear unbiased estimators. Thus in the finite population sampling case, it is relevant to know if some y_i's belonged to the same unit repeated, or to two different units. This fact was first noticed by Des Raj and Khamis (1958) and Basu (1958) who observed that for a simple random sample of size n taken with replacement, the mean of the effective sample, viz., $\frac{1}{v_s} \Sigma_1^{v_s} y_i'$, y_i''s being the y-values corresponding to the distinct units, is better than the conventional sample mean $\bar{y}$ based on all units in the sample. This estimator has coefficients attached to y_i which unlike the ω_i's mentioned above, depend not only on v_s, the effective sample size, but also on the sample selected till $(i-1)$ draws. This necessitated a general definition of linear estimators and Horvitz and Thompson (1952) were the first to formulate the problem in accordance with a unified approach and defined three classes of linear estimators for the population total $T = \Sigma_i^N Y_i$ which are as follows

(i) $\hat{T}_1 = \Sigma_{i=1}^{n} \alpha_i y_i$, where α_i is a constant to be used as a weight for the unit selected in the ith draw, i = 1, 2, ..., n.

(ii) $\hat{T}_2 = \Sigma_{i=1}^{n} \beta_i y_i$, where β_i is a constant to be used as weight for the ith unit whenever it is selected for the sample and

(iii) $\hat{T}_3 = \gamma_{s_n} (\Sigma y_i)_{s_n}$, where γ_{s_n} is a constant to be used as weight whenever the s_nth sample is selected.

Horvitz and Thompson, in their 1952 paper, considered the T_2–class of estimators and showed that the only unbiased linear estimator possible in this subclass is the one with $\beta_i = 1/\pi_i$, where π_i is the probability of inclusion of ith unit for the sampling design. Thus we have the 'celebrated' Horvitz-Thompson (HT) estimator which is the best in this subclass, given by

$$\hat{Y}_{HT} = \sum_{i \in s} y_i / \pi_i \tag{1.1}$$

for the estimation of the population total $Y = \Sigma_{i=1}^{N} Y_i$.

Horvitz and Thompson, indeed recognize that this is the "only unbiased linear estimator possible in the subclass under consideration and hence is 'best' for that subclass". It is easy to derive that

$$V(\hat{Y}_{HT}) = \sum_{i=1}^{N} \left(\frac{1}{\pi_i} - 1 \right) Y_i^2 + \sum_{i \neq j}^{N} \sum^{N} \left(\frac{\pi_{ij}}{\pi_i \pi_j} - 1 \right) Y_i Y_j \tag{1.2}$$

Horvitz and Thompson (HT) have also given an unbiased estimator of the above variance expressed as

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

One notes from their 1952 paper the following

"... It is the opinion of the authors that the techniques suggested by this paper may be of greatest utility in specialized enquiries where the characteristics under measurement are few and related, or where selection with unequal probability arises naturally. The estimator (6) (of their paper, i.e. $\Sigma y_i / \pi_i$) from a computational point of view is at a serious disadvantage when compared with self-weighting estimators ...".

They also noted that "... when an unbiased estimator of high precision and an ubiased estimator of its variance are required, however, the sampling system employing unequal probabilities, with the selection of two or more units at each stage of sampling, may be particularly appropriate ..."

HT observed that when $\pi_i = nY_i / \Sigma_1^N Y_i$, $\hat{Y}_{HT}$ will have zero variance and the sampling will be optimum. When auxiliary information on a characteristic x is available taking values X_i on units U_i, i = 1, 2, ..., N such that X_i's can be assumed to be roughly proportional to Y_i's, then $\pi_i = nP_i$, with $P_i = X_i/X$ would be near-optimum, where $X = \Sigma_{i=1}^N X_i$.

Thus for the case of PPS selection WOR for n = 2, one would like to find "working probabilities, "say Q_i associated with units U_i, so that

$$\begin{aligned}\pi_i &= Q_i + \sum_{j \neq i}^{N} Q_i \frac{Q_j}{1-Q_j} \\ &= Q_i\left(1 + Q - \frac{Q_i}{1-Q_i}\right), \text{ with } Q = \sum_{i=1}^{N} \frac{Q_i}{1-Q_i} \\ &= 2P_i \end{aligned} \tag{1.4}$$

HT recommend using

$$2Q_i(1-Q_i) = 2P_i$$

for solving Q_i, while Yates and Grundy (1953) in their paper consider

$$\pi_i = 2P_i = Q_i\left(1 + Q - \frac{Q_i}{1-Q_i}\right)$$

and give a first approximation as

$$Q_i^{(1)} = \frac{2P_i}{1 + P - \frac{P_i}{1-P_i}}, \text{ with } P = \sum_{i=1}^{N} \frac{P_i}{1-P_i}$$

and so on.

We remark here that Yates had considered the estimator

$$\hat{Y}_U = \frac{1}{2}\left(\frac{y_i}{p_i'} + \frac{y_j}{p_j'}\right) \tag{1.5}$$

where $2p_i' = p_i\left(1 + \Sigma_{j \neq i}^N \frac{p_j}{1-p_j}\right)$

for *ppswor* selection of n = 2 units. He mentions in Yates and Grundy (1953) "The present investigation was originally undertaken by the first author (Yates) with the object of determining in more detail the errors likely to result from this procedure of selection and estimation" ... . "It was only after the paper had been prepared and submitted for publication that a copy of a recent paper by Horvitz and Thompson (1952) on the same subject became available ..." Thus (1.5) is nothing but $\hat{Y}_{HT}$ for n = 2.

A forerunner to these results is the soluation given by Narain (1951) from India, wherein Q_i are obtained from

$$Q_i\left(1+Q-\frac{Q_i}{1-Q_i}\right)=\pi_i=2P_i \tag{1.6}$$

In this paper, Narain (1951) has brought out "certain features novel to sampling wor."

In Yates and Grundy (1953) it is acknowledged thus : ...Narain (1951) has given an alternative solution of which we are unaware until after this paper had been sent to press...". Perhaps one should call $\Sigma y_i/\pi_i$ as the "Narain-Yates-Horvitz-Thompson" estimator! JNK Rao (1999) calls it Narain-Horvitz-Thompson estimator.

We note from Yates and Grundy (1953)'s paper the following

"... Although Horvitz and Thompson's paper deals with much the same problem, the conclusions reached by the first author (Yates) differed from theirs in many respects..." . "In one respect, Horvitz and Thompson took the matter further in that they gave an unbiased estimator of the error variances. This however, proved on examination to be unsatisfactory... (negative for some samples). This fact stimulated the second autor (Grundy) of the present (revised) paper to search for a better unbiased estimator which is included in the revised version ...". However, their variance estimator also turns out to be negative but less often than HT's.

The variance expression was rewritten in the form

$$V(\hat{Y}_{HT})=\sum\sum_{i<j=1}^{N}(\pi_i\pi_j-\pi_{ij})\left(\frac{Y_i}{\pi_i}-\frac{Y_j}{\pi_j}\right)^2 \tag{1.7}$$

for any fixed sample size (n) design, in Yates and Grundy (1953) besides Sen (1953). Thus we have the well-known form of the variance estimator (Sen-Yates-Grundy estimator or perhaps Sen-Grundy estimator) given by

$$\hat{V}_{SYG}(\hat{Y}_{HT})=\sum\sum_{i<j\in s}\frac{(\pi_i\pi_j-\pi_{ij})}{\pi_{ij}}\left(\frac{y_i}{\pi_i}-\frac{y_j}{\pi_j}\right)^2 \tag{1.8}$$

Concerning the non-negativity of (1.8), while Yates and Grundy (1953) felt that "this appears to be the case when the usual method of selection is employed", Sen (1953) from India quite independently, went on to obtain sampling schemes for which (1.8) is always non-negative.

The expression for the variance of $\hat{Y}_{HT}$ also suggests that if $\pi_i \propto Y_i$, the estimator would be most precise, a fact also noted by Horvitz and Thompson. Since Y_i's are unknown, if related auxiliary information on a characteristic x is available, then a suitable choice for a design would be one for which $\pi_i \propto x_i$.

2. Optimality of HT Estimator

Godambe (1955) established that there does not exist a umvue in the general class of sampling designs. He has used the Superpopulation concept introduced by Cochran (1939, 1946) and established that under the class of distributions satisfying $E(Y_i \mid X_i) \propto X_i$, $V(Y_i \mid X_i) \propto X_i^2$ and $C(Y_i, Y_j \mid X_i, X_j) = 0$, an optimum strategy for which (i) $\pi_i \propto X_i$, (ii) $v_s = v \; \forall \, s$ with $p_s > 0$ and (iii) $\hat{Y} = \hat{Y}_{HT}$ exists which has minimum expected variance.

This result opened up the construction of π PS sampling schemes which insisted on non-negative variance estimation and stability of variance estimator as well. About 80 such schemes are available in literature. Other optimality properties of the HT estimator include admissibility, hyper-admissibility etc.

It may be pointed out that whenever $\hat{V}(\hat{Y}_{HT})$ takes negative values, one could consider the biased truncated estimator

$$\begin{aligned} \hat{V}^* &= \hat{V}(\hat{Y}_{HT}) \text{ if } \hat{V}(\hat{Y}_{HT}) > 0 \\ &= 0 \qquad\quad \text{if } \hat{V}(\hat{Y}_{HT}) < 0 \end{aligned} \tag{2.1}$$

If $\hat{V}_{HT}(\hat{Y}_{HT})$, for example, is negative, one could also suggest

$$\hat{V}' = \frac{\hat{V}_{HT}(\hat{Y}_{HT})}{\hat{V}_{HT}(\hat{X}_{HT})} V_{HT}(\hat{X}_{HT}) \tag{2.2}$$

whenever $V_{HT}(\hat{X}_{HT})$ is non-zero. Here since y and x are related it is expected that both the numerator and denominator tend to be of the same sign.

As mentioned earlier, for a particular class of super population models it was shown that the strategy consisting of πPS sampling scheme and the corresponding Horvitz-Thompson estimator is optimum, in the sense of minimum expected variance under the model. However, there are alternative techniques of estimating the population total such as ratio method of estimation giving

$$\hat{Y}_{Ratio} = \frac{\sum_{i \in s} y_i}{\sum_{i \in s} x_i} X \tag{2.3}$$

This being biased for the population total Y for simple random sampling, an unbiased ratio-type estimator is obtained utilizing the Midzuno-Sen scheme. It is easy to calculate the probability of inclusion π_i' for the Midzuno-Sen scheme which is given by $\pi_i' = \frac{n-1}{N-1} + \frac{N-n}{N-1}\frac{x_i}{X}$. Then, one could immediately consider a corresponding unbiased strategy

$$\mathcal{H}\left(\text{Midzuno-Sen scheme}, \hat{Y}_{HT}' = \sum_{i=1}^{n} y_i/\pi_i'\right)$$

for the estimation of Y. Note that this is a non-πPS sampling scheme and comparisons have shown the superiority of πPS strategy. In this context it is interesting to note that the competing strategies, viz., Symmetrized Des Raj (Murthy's) strategy or Rao-Hartley-Cochran (RHC) strategy do not perform so well compared to πPS strategy with $\hat{Y}_{HT}$, when the super population model parameter g is close to 2 (Here $\mathcal{V}(Y_i \mid X_i)$ is assumed to be proportional to X_i^g). With its optimality properties, it is but natural to consider Generalized Ratio Estimator (Hajek's)

$$\frac{\sum_{i \in s} y_i/\pi_i}{\sum_{i \in s} x_i/\pi_i} X \text{ and} \tag{2.4}$$

the Generalized Regression Estimator (GREG), viz.

$$\sum_{i \in s} \frac{y_i}{\pi_i} + \hat{\beta}\left(X - \sum_{i \in s} \frac{x_i}{\pi_i}\right) \tag{2.5}$$

in the theory of small area estimation, especially the latter.

When certain model assumptions are satisfied and when design-unbiasedness is not demanded, it was demonstrated by Royall (1970) that an estimator like

$$\sum_{i \in s} y_i + \left(\frac{1}{n}\sum_{i \in s}\frac{y_i}{x_i}\right)\left(X - \sum_{i \in s} x_i\right) \tag{2.6}$$

could perform better than the Horvitz-Thompson estimator $\hat{Y}_{HT}$. However this theory depends very much on the model assumptions. Rao (1971) considered the set of parameter values $(\pi_1 \gamma_{i1}, \pi_2 \gamma_{i2}, \ldots, \pi_N \gamma_{iN})$ obtained by permutations of

$(\gamma_1, \gamma_2, \ldots, \gamma_N)$ keeping $(\pi_1, \pi_2, \ldots, \pi_N)$ fixed where $\gamma_i = \frac{y_i}{\pi_i}$. For fixed sample size designs, C.R. Rao then demonstrated the 'optimality' of the HT estimator $\hat{\bar{Y}}_{HT} = \Sigma_{i \in s} y_i \mid N\pi_i$ in the class of linear design-unbiased estimators of the population mean $\bar{Y}$. This important property of HT estimator was later on extended to wider classes of estimators by several others.

It may be pointed out that the Politz-Simmons (1949) technique for the problem of 'not-at-homes' in an interview survey was an off-shoot of the Horvitz–Thompson (1952) paper as observed by them.

3. Remarks

In this section we shall briefly mention some recent lines of research where Horvitz-Thompson estimator plays an important role.

The problem of studying a three dimensional physical object from random two-dimensional plane sections or projections and the subsequent techniques of estimation of geometrical parameters, namely, volume, surface area, total curvature etc. is the subject matter of Stereology. Baddeley (1993) regards Modern Stereology as "sampling theory for spatial processes". Baddeley (1993) obtains 'edge-correction' estimators for characteristics of point processes which take the form of a ratio of two unbiased Horvitz-Thompson estimators. He remarks that "the analogy with Horvitz-Thompson is not strong enough to improve the estimation of variances". There is further scope for studying this problem.

Notwithstanding the optimality properties, the Horvitz-Thompson estimator is not robust against outliers. However, Hulliger (1995) expresses the HT estimator as a least squares (LS) functional of an estimate of the population distribution function while the asumption of proportionality between y_i and x_i is utilized in the LS-functional.

Let

$$F_s(r, t) = \sum_{i \in s} \frac{1}{\pi_i} I\{x_i \leq r\} I\{y_i \leq t\} \Big/ \sum_{i \in s} \frac{1}{\pi_i} \tag{3.1}$$

Assume that Y_i's are independent with expectation βx_i and variance $\sigma^2 x_i$. Under the model, the LS-estimator $\beta_{LS}(F_s)$ of β w.r.t. sampling distribution F_s of $(x_i, y_i), i \in s$ minimizes

$$\int (y - \beta x)^2 / x \, dF_s(x, y) \tag{3.2}$$

or equivalently solves

$$\sum_{i \in s} \frac{1}{\pi_i} \left(\frac{y_i - \beta x_i}{\sqrt{x_i}} \right) \frac{x_i}{\sqrt{x_i}} = 0 \tag{3.3}$$

For πPS sampling design, the HT estimator

$$\hat{Y}_{HT} = X\beta_{LS}(F_s) \tag{3.4}$$

where $\beta_{LS}(F_s)$ the LS-estimator of π is defined by (3.3) is given by

$$\beta_{LS}(F_s) = \frac{\Sigma_{i \in s} y_i / \pi_i}{\Sigma_{i \in s} x_i / \pi_i}$$

The estimating equation (3.3) depends on the residuals $y_i - \beta x_i$ and x_i. Writing $\frac{(y_i - \beta x_i)}{x_i^{1/2}}$ as $r_i'(\beta)$, the standardized residuals and $x_i / x_i^{1/2}$ as x_i', we observe that (3.3) can be written as

$$\sum_{i \in s} \frac{1}{\pi_i} r_i'(\beta) x_i' = 0$$

Defining $\beta(F_s, \eta)$ as a solution of the equation

$$\sum_{i \in s} \frac{1}{\pi_i} \eta(x_i', r_i'(\beta)) x_i' = 0 \tag{3.5}$$

Hulliger (1995) defines $\hat{\bar{Y}}_{RHT} = \beta(F_s, \eta)\bar{X}$ as the Robustified Horvitz-Thompson estimator. The problem of estimation of variance is also considered in Huliger (1995).

Under a general super population model with $v(Y_i) = \sigma^2 v(x_i)$, we have (3.3) as the estimating equation where in now

$$r_i'(\beta) = \frac{y_i - \beta x_i}{\sqrt{v(x_i)}} \text{ and } x_i' = \frac{x_i}{\sqrt{v(x_i)}}$$

In this context it will be interesting to see how the usual competitor to HT estimator, viz. Rao-Hartley-Cochran (RHC) estimator could be robustified. To save length and to maintain the theme of the paper, we shall only outline the procedure and expand on this and related results elsewhere.

Form n random subgroups of sizes $N_1, N_2, \ldots, N_n$. Draw one unit from each group with probability proportional to size. Then we have

$$\hat{\bar{Y}}_{RHC} = \frac{1}{N} \sum_{i=1}^{n} \frac{y_i}{x_i / \Sigma x_i} \tag{3.6}$$

where Σx_i is the total measure of size for group i.

Mimicking the above results, we have

$$F_s^*(r,t) = \sum_{i=1}^{n} \frac{I(x_i \le r) I(y_i \le t)}{p_i/\Sigma_1^n p_i} \Big/ \sum_{i=1}^{n} \frac{1}{p_i/\Sigma_1^n p_i}$$

Here we write the estimating equation as

$$\sum \frac{1}{(p_i/\Sigma_1^n p_i)} \left(\frac{y_i - \beta x_i}{\sqrt{x_i}} \right) \frac{x_i}{\sqrt{x_i}} = 0 \tag{3.7}$$

for the super population model with $v(Y_i) \propto x_i$, whose solution gives $\hat{\beta}_{LS}$.

As before define $\beta^*(F_s, \eta)$ as a solution of the equation

$$\sum_{i=1}^{n} \frac{1}{(p_i/\Sigma p_i)} \eta(x_i', r_i'(\beta)) x_i' = 0 \tag{3.8}$$

We then have $\hat{\bar{Y}}_{RRHC} = \beta^*(F_s, \eta)\bar{X}$ as the Robustified RHC estimator. Arguing similarly the Robustified Murthy estimator may be written as

$$\hat{Y}_{RSDR} = \beta'(F_s, \eta)\bar{X}$$

where β' is a solution of

$$\sum_{i=1}^{n} \frac{1}{(p(s \mid i)/p(s))} \eta(x_i, r_i'(\beta)) x_i' = 0$$

where the symbols have the usual meaning as in Murthy's estimator.

Next we consider Basu's (1971) frivolous circus example which gives undue inclusion probabilities to the elephants (the units here) which leads to Horvitz-Thompson estimator being absurd. He is not insisting the mathematical property of unbiasedness of an estimator but concerns himself with the "hard-to-define property of face validity" of an estimate. In a certain sense, he seeks the 'consistency' of the estimate. Ghosh (1992) establishes that the HT estimator is consistent and with probability tending to 1, it approximates a reasonable design free estimate and thus has some 'face validity'. A proper mathematical definition of 'face validity' and its role in the choice of estimators is to be further established.

Cassel and Sarndal (1974) evaluated some sampling strategies for finite populations using a continuous variable frame work. When a sample of points is to be selected from a geographical region, namely a district or a forest area, the finite population approach for estimation of parameters is not applicable. In this case of point sampling from a continuous universe, Cordy (1993) developed a theory for Horvitz-Thompson estimator and its variance estimation.

An ordered n-tuple $(x_1, x_2, \ldots, x_n)$ of elements from a continuous population U is an ordered sample s of size n and a collection of s is the sample space $\mathcal{S}$. A sampling design on $\mathcal{S}$ is a probability measure defined through a p.d.f. of n random variables $X_1, \ldots X_2, \ldots, X_n$ with values in U, defined on $\mathcal{S}$ with certain conditions. Let f_i be the marginal distribution of X_i. For $x \in U$, define

$$\pi(x) = \sum_{i=1}^{n} f_i(x)$$

This defines the 'inclusion density function' on U. Similarly, denoting by f_{ij}, the marginal density of X_i and X_j, for $x, x' \in U$, we define the 'joint inclusion density function'

$$\pi(x, x') = \sum \sum_{i=1 \neq j}^{n} f_{ij}(x, x')$$

For estimating the total $Y = \int_U y(x)\,dx$, Cordy (1993) considers the Horvitz-Thompson estimator

$$\hat{Y} = \sum_{x \in s} y(x)/\pi(x), \text{ for } s \in \mathcal{S}$$

We shall quote an interesting theorem due to Cordy, where in the first terms of the formulae are a departure from the familiar finite population set up. *Theorem* (Cordy (1993)): Suppose that the function y is bounded, $\pi(x) > 0$, for each $x \in U$ and $\int_U (1/\pi(x))dx < \infty$. Then

$$V(\hat{Y}_{HT}) = \int_U ((y(x))^2 / \pi(x))\,dx + \int_U \int_U y(x)y(x')\{(\pi(x, x') - \pi(x)\pi(x')) / \pi(x)\,\pi(x')\}dx\,dx'$$

Furthermore, if $\pi(x, x') > 0$ for all $x, x' \in U$, then

$$\hat{V}(\hat{Y}_{HT}) = \sum_{x \in s} (y(x)/\pi(x))^2 + \sum_{x \in s} \sum_{\substack{x' \in s \\ x \neq x'}} y(x)\,y(x')\{(\pi(x, x') - \pi(x)\pi(x'))/\pi(x, x')\pi(x)\pi(x')\}$$

Padmawar (1996) applied this theory to certain other estimators as well and compared them with HT estimator.

We shall now briefly discuss a related problem in Environmental Sampling Protocol wherein an integration approach to HT estimator is suggested. Consider a population of N units in a planar region under study. Let y_i be the number of marks (citings) for the ith unit in the context of environment sampling, so that $Y = \Sigma y_i$, the population total is the parameter of interest. Consider a base line (without loss of generality, the interval (0, 1)). Let x represents the projection of the selected point on to the base line. Then the HT estimator of Y is defined as

$$\hat{Y}(x) = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_{P_i}(s)$$

where P_i is the inclusion set for the ith unit, viz. a suitable interval contained in the base line and I is the indicator function. In environmental protocols, a replicated sampling of n points is taken and the replicated method gives rise to n HT estimators, $\hat{Y}(x_i), i = 1, 2, \ldots n$. Using the fact that

$$\pi_i = \int_0^1 I_{P_i}(x)dx$$

we express $\hat{Y}$ as

$$\hat{Y} = \int_0^1 \hat{y}(x)dx$$

thereby reducing the problem of estimation to a problem of integration, where the integrand is an unknown function evaluated at the points $(x_1, x_2 \ldots x_n)$. Thus the methodology for numerical quadrature could be used and choice of $(x_1, x_2, \ldots x_n)$ be made to construct an estimator of the type

$$\bar{t} = \bar{t}(x_1, x_2, \ldots, x_n) = \frac{1}{n}\sum_{i=1}^{n} \hat{Y}(x_i)$$

Barabesi (2003) describes various Monte Carlo Integration approaches for this problem.

It should be pointed out here that in the additional remarks made by D. Basu during the discussion on his paper "An essay on the logical foundations of survey sampling, Part I" on April 3, 1970 it is mentioned as follows

"... An analogy was drawn between the problem of estimating the population total ΣY_j and the classical problem of numerical integration. In the latter the problem is to 'estimate' the value of the integral $\int_a^b Y(u)du$ by 'surveying' the function Y(u) at a number of 'selected points' $u_1, u_2, \ldots, u_n$. Which points to select and how many of them, are the problems of 'design'.

Which integration formula to use and how to assess the 'error' of estimation, are problems of 'analysis'. True, it is possible to set up a statistical theory of numerical integration by forcing an element of randomness in the choice of the points. ..."

Finally, in the context of the Horvitz-Thompson estimator, we shall refer to some results which use matrix algebra. Dol *et al.* (1996) show the convenience of the use of a number of matrix algebra results. They derive conditions, under which $V(\hat{Y}_{HT})$ goes to zero asymptotically and hence sufficient conditions for the consistency and the rate of convergence of the HT estimator are given. These conditions are based on matrix inequalities that are employed to obtain a bound for the variance of $\hat{Y}_{HT}$. Using matrix algebra results, Taga (1993) gave a generalization of the HT strategy. Application of matrix results makes the algebra elegant as seen in several recent works (see for example, Valliant *et al.* (2000)).

4. Conclusion

In this paper we have traced rather critically the history relating to the Horvitz-Thompson estimator and reviewed briefly some of the important optimality properties. While improvisations and improvements in the current theoretical work are going on, there are new areas where researchers have taken note of the HT estimator. We have briefly outlined some of these developments here and it is clear that the Horvitz-Thompson estimator would continue to play an important role in sampling theory and applications.

REFERENCES

Baddeley, A.J. (1995). Stereology and survey sampling theory. *Bull. Int. Stat. Inst.*, **55(2)** 435-449.

Barabesi, L. (2003). A Monte Carlo Integration approach to Horvitz-Thompson estimation in replicated environmental designs. *Metron*, **61(3)** 355-374.

Basu, D. (1958). On sampling with and without replacement. *Sankhya*, **20**, 287-294.

... (1971). An essay on the logical foundations of survey sampling-part one. In : V.P. Godambe and D.A. Sprott (Eds.), *Foundations of Statistical Inference*, Toronto : Holt, Rinehart & Winston, 203-242.

Cassel, C.M. and Sarndal, C.E. (1974). Evaluation of some sampling strategies for finite populations using a continuous variable frame work. *Comm. Stat.*, **3**, 373-390.

Cochran, W.G. (1939). The use of analysis of variance in enumeration by sampling, *J. Amer. Statist. Assoc.*, **34**, 492-510.

... (1946). Relative accuracy of systematic and stratified random samples for a certain class of populations. *Ann. Math. Statist.*, **17**, 164-177.

Cordy, C.B. (1993). An extension of the Horvitz-Thompson theorem to point sampling from a continuous universe. *Statist. Probab. Lett.*, **18**, 353-362.

Des Raj and Khamis, S.H. (1958). Some remarks on sampling with replacement. *Ann. Math. Statist.*, **29**, 550-557.

Dol, W., Steerneman, T. and Wansbeek, T. (1996). Matrix algebra and sampling theory : The case of the Horvitz-Thompson estimator. *Linear Alg. Appl.*, **237/238**, 225-238.

Ghosh, J.K. (1992). The Horvitz-Thompson estimate and Basu's Circus revisited, *Bayesian Analysis in Statistics and Econometrics* (Eds.) P.K. Goel and N.S. Iyengar, 225-228.

Godame, V.P. (1995). A unified theory of sampling from finite populations. *J. Roy. Statist. Soc.*, **B17**, 269-278.

Horvitz, D.G. and Thompson, D.J. (1952). A generalisation of sampling without replacement from finite-universe. *J. Amer. Statist. Assoc.*, **47**, 663-685.

Hulliger, B. (1995). Outlier Robust Horvitz-Thompson estimator. *Survey Methodology*, **21**, 79-87.

Narain, R.D. (1951). On sampling without replacement with varying probabilities. *J. Ind. Soc. Agril. Statist.*, **3**, 169-174.

Padmawar, V.R. (1996). Rao-Hartley-Cochran strategy in survey sampling of continuous populations. *Sankhya*, **58B**, 90-104.

Politz, A.N. and Simmons, W.R. (1949). An attempt to get the "not-at-homes" into the sample without call backs. *J. Amer. Statist. Assoc.*, **44**, 9-31.

Rao, C.R. (1971). Some aspects of statistical inference in problems of sampling from finite populations. In : V.P. Godame and D.A. Sprott (: ds.), *Foundations of Statistical Inference*, Toronto : Holt, Rinehart & Winston, 177-202.

Rao, J.N.K. (1999). Some current trends in sample survey theory and methods. *Sankhya*, **B61**, 1-57.

Royall, R.M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, **57**, 377-387.

Sen, A.R. (1953). On the estimate of variance in sampling with varying probabilities. *J. Ind. Soc. Agril. Statist.*, **5**, 119-127.

Taga, Y. (1993). Generalization of HT strategy and its application, Yokohama *Math. Jour.*, **40**, 163-173.

Valliant, R., Dorfman, A.H. and Royal, R.H. (2000). *Finite Population Sampling and Inference : A Predication Approach.* Wiley.

Yates, F. and Grundy, P.M. (1953). Selection without replacement from within strata with probability proportional to size. *J. Roy. Statist. Soc.*, **B15**, 253-261.