© Journal of the American Statistical Association
December 1971, Volume 66, Number 336
Theory and Methods Section

πPS Sampling Designs and the Horvitz-Thompson Estimator

T. J. RAO*

Using the criterion of minimum expected variance and the Horvitz-Thompson estimator, we study various πPS (π_i, the probability of inclusion of the ith unit, Proportional to Size) strategies and make a comparison between these strategies under a general super population model. This discussion further motivates the search for a class of designs which are best suited for the use of the Horvitz-Thompson estimator. A new class of such designs is obtained.

1. INTRODUCTION

Hansen and Hurwitz [6] demonstrated the profitability of selecting sampling units with probability proportional to size of the unit and indicated methods of determining the probabilities of selection which minimize the variance of the sample estimate at a fixed cost. They also showed [7] that sampling with probability proportional to the square root of size is more efficient than sampling with probability proportional to size under certain conditions. Later, in 1952 Horvitz and Thompson [11] first recognized the need for dealing systematically with the theory of sampling from finite populations and, besides formulating the theory neatly, they defined three classes of estimators. Subsequently, in 1955 Godambe [3] proposed a unified theory of sampling from finite populations with a view to discussing the fundamental problems of sampling within this framework and also formulated the definition of linearity with a general theory of sampling.

Godambe [3] established that for any sampling design there does not exist a uniformly minimum variance unbiased estimator of the population total in the class of all linear unbiased estimators (barring certain exceptions, characterized later). However it was first shown by Cochran in 1946 [2] that whenever auxiliary information on a characteristic $\mathfrak{X}$ which takes values X_i on the unit U_i, $i=1, 2, \cdots, N$ is available closely related to the characteristic $\mathcal{Y}$ under study, taking values Y_i on U_i, $i=1, 2, \cdots, N$, it is possible to use this information to set up a criterion of optimality. Thus, according to this 'super population concept', $Y=(Y_1, Y_2, \cdots, Y_N)$ is assumed to be a realization of an N-length random vector with distribution θ depending on $X=(X_1, X_2, \cdots, X_N)$ and some unknown parameters. We explicitly formulate our general model Θ_g thus:

$$\left.\begin{aligned} \mathcal{E}_\theta(Y_i \mid X_i) &= aX_i \\ \mathcal{V}_\theta(Y_i \mid X_i) &= \sigma^2 X_i^g \\ \mathcal{C}_\theta(Y_i, Y_j \mid X_i, X_j) &= 0 \end{aligned}\right\} \tag{1.1}$$

where the script letters $\mathcal{E}$, $\mathcal{V}$ and $\mathcal{C}$ denote the conditional expectation, variance and covariance given X_i's. The expected variance $\int \text{Var}(H)d\theta$ of the sampling strategy H (sampling design together with an estimator being called a 'strategy') is now minimized over the class of all equi-cost strategies and a strategy that minimizes this expected variance is called a 'Θ_g-optimum' strategy.

Using this concept Godambe [3] proved that under the particular model Θ_2 ((1.1) with $g=2$), there exists a Θ_2-optimum strategy for which

a. The inclusion probability of the ith unit, π_i is proportional to the value X_i taken by the auxiliary characteristic on that unit,
b. Every sample has n distinct units and
c. The estimator used is the corresponding Horvitz-Thompson [11] estimator

$$\hat{Y}_{HT} = \sum_{i\in s} (Y_i/\pi_i) \tag{1.2}$$

for the estimation of the population total $Y=\sum_{i=1}^{N} Y_i$ where the symbol $\sum_{i\in s}$ indicates that the summation is over the distinct units of the sample s,

in the class of all unbiased strategies with n distinct units. Later Hanurav [8] in 1962 obtained a class of optimal sampling designs best suited for the use of the Horvitz-Thompson estimator and termed [9] them as πPS (π_i's Proportional to Size) sampling designs. Using the criterion of minimum expected variance and the Horvitz-Thompson estimator, we study in this article various πPS designs and present a comparison between these designs under the general super population model (1.1). This discussion then leads to the investigation of the optimum choice of the measure of size to employ when sampling with probability proportional to modified size in conjunction with the use of Horvitz-Thompson estimator.

* T. J. Rao is lecturer, Statistical Laboratory, Department of Mathematics, University of Manchester, Manchester M13 9PL, England. The author is on leave from the Indian Statistical Institute. He is grateful to the referee for his comments which were useful in restructuring the article and to Miss Kate Cross for her excellent typing.

2. COMPARISON BETWEEN πPS STRATEGIES

Considering the Horvitz-Thompson [11] estimator $\hat{Y}_{HT} = \sum_{i\in s} Y_i/\pi_i$ defined in (1.2) we have

$$\begin{aligned}\mathrm{Var}(\hat{Y}_{HT}) &= E(\hat{Y}^2_{HT}) - Y^2\\ &= \sum_{i=1}^{N} (1/\pi_i - 1) Y_i^2\\ &\quad + \sum_{i\neq j}^{N} \sum^{N} (\pi_{ij}/\pi_i\pi_j - 1) Y_i Y_j \qquad (2.1)\end{aligned}$$

where π_i is the probability of inclusion of the ith unit in the sample and π_{ij} is the probability of joint inclusion of the ith and jth units in the sample. Further, under the model Θ_g of (1.1) we have

$$\begin{aligned}&\mathcal{E}_{\theta(g)} \mathrm{Var}\{(\textstyle\sum_{i\in s} Y_i/\pi_i)\}\\ &= \sum_{i=1}^{N} (1/\pi_i - 1)\mathcal{E}(Y_i^2 \mid X_i)\\ &\quad + \sum_{i\neq j}^{N} \sum^{N}(\pi_{ij}/\pi_i\pi_j - 1)\mathcal{E}(Y_iY_j \mid X_i, X_j)\\ &= \sum_{i=1}^{N} (1/\pi_i - 1)(\sigma^2 X_i^g + a^2 X_i^2)\\ &\quad + \sum_{i\neq j}^{N} \sum^{N}(\pi_{ij}/\pi_i\pi_j - 1)a^2 X_i X_j\\ &= \sigma^2 \sum_{i=1}^{N} (1/\pi_i - 1)X_i^g + a^2 \mathrm{Var}\{\textstyle\sum_{i\in s}(X_i/\pi_i)\}. \qquad (2.2)\end{aligned}$$

The minimum value of (2.2) for Godambe's Θ_g-optimum strategy is given by

$$\sigma^2 \sum_{i=1}^{N} (1/nP_i - 1)X_i^2 \quad \text{where} \quad P_i = X_i/\sum_{i=1}^{N} X_i$$

since the second term

$$\mathrm{Var}\{\textstyle\sum_{i\in s}(X_i/\pi_i)\} = \mathrm{Var}\{\textstyle\sum_{i\in s}(X_i/nP_i)\}$$

vanishes.

Let the "effective sample size" $\nu(s)$ be defined as the number of distinct units in a sample s. Assuming that the cost of drawing and inspecting the sample s is proportional to the effective sample size $\nu(s)$, it is reasonable to compare strategies for which the expected value of $\nu(s)$ is a given value and this would mean that the expected cost of sampling is fixed beforehand. Notice that

$$E(\nu(s)) = \sum_{s\in S} \nu(s)p_s = \sum_{i=1}^{N} \pi_i = \nu_0,$$

say, where p_s is the probability attached to the sample s such that p_s summed over the collection S of all samples is unity.

Let X_i^α be the generalized measure of size, where α is a real number and let ν_0 be the expected cost. We have from (2.2) when $\pi_i \propto X_i^\alpha$ that

$$\begin{aligned}&\mathcal{E}_{\theta(g,\alpha)} \mathrm{Var}\{\textstyle\sum_{i\in s} (Y_i/\pi_i)\}\\ &= \sigma^2\left[\frac{\sum_{i=1}^{N} X_i^{g-\alpha} \sum_{i=1}^{N} X_i^\alpha}{\nu_0} - \sum_{i=1}^{N} X_i^g\right] + \Delta(\alpha) \qquad (2.3)\end{aligned}$$

where

$$\Delta(\alpha) = a^2 \mathrm{Var}\{\textstyle\sum_{i\in s} (X_i/\pi_i)\}$$

with $\pi_i \propto X_i^\alpha$. Notice that when $\alpha = 1$, we have πPX sampling and $\sum_{i\in s} (X_i/\pi_i)$ is equal to a constant so that its variance and, hence, $\Delta(1)$ vanishes. Thus

$$\begin{aligned}&\mathcal{E}_{\theta(g,1)} \mathrm{Var}\{\textstyle\sum_{i\in s} (Y_i/\pi_i)\}\\ &= \sigma^2\left[\frac{\sum_{i=1}^{N} X_i^{g-1} \sum_{i=1}^{N} X_i}{\nu_0} - \sum_{i=1}^{N} X_i^g\right]. \qquad (2.4)\end{aligned}$$

It can be shown from (2.3) that

$$\phi(\alpha) = \sum_{i=1}^{N} X_i^{g-\alpha}\sum_{i=1}^{N} X_i^\alpha \geq (\textstyle\sum_{i=1}^{N} X_i^{g/2})^2$$

so that $\alpha = g/2$ minimizes $\phi(\alpha)$ and hence the first term of (2.3). Thus we have

$$\begin{aligned}&\mathcal{E}_{\theta(g,g/2)} \mathrm{Var}\{\textstyle\sum_{i\in s} (Y_i/\pi_i)\}\\ &= \sigma^2\left[\frac{(\sum_{i=1}^{N} X_i^{g/2})^2}{\nu_0} - \sum_{i=1}^{N} X_i^g\right] + \Delta(g/2) \qquad (2.5)\end{aligned}$$

and the difference between (2.4) and (2.5) is given by

$$\begin{aligned}(\sigma^2/\nu_0)[\textstyle\sum_{i=1}^{N} X_i^{g-1} \sum_{i=1}^{N} X_i - (\sum_{i=1}^{N} X_i^{g/2})^2]\\ - \Delta(g/2). \qquad (2.6)\end{aligned}$$

Observing that the first term of (2.6) is positive, we note that when $\Delta(g/2)$ is small enough, the sampling schemes where $\pi_i \propto X_i^{g/2}$ would fare better than those for which $\pi_i \propto X_i$ (cf. [7]).

Let us denote (2.4) by $\mathcal{E}_{\pi PS}$ and (2.3) by $\mathcal{E}_{\pi PMS\alpha}$ where πPMSα stands for π_i's Proportional to Modified Size X_i^α. Comparison of (2.3) and (2.4) leads to

Lemma 2.1.

$$\frac{\mathcal{E}_{\pi PMS\alpha} - \mathcal{E}_{\pi PS}}{\sigma^2} > \frac{1}{\nu_0} f(\alpha) \qquad (2.7)$$

where

$$f(\alpha) = \sum_{i=1}^{N} X_i^{g-\alpha} \sum_{i=1}^{N} X_i^\alpha - \sum_{i=1}^{N} X_i^{g-1} \sum_{i=1}^{N} X_i.$$

The proof is omitted.

We next state a result due to Callebaut [1] in the following lemma which is useful in this context.

Lemma 2.2. For $a = (a_1, a_2, \cdots, a_N)$ and $b = (b_1, b_2, \cdots, b_N)$, positive vectors which are not proportional, the expression

$$(\textstyle\sum_{i=1}^{N} a_i^{y+z} b_i^{y-z})(\sum_{i=1}^{N} a_i^{y-z} b_i^{y+z})$$

increases with increasing $|z|$ for any real number y.

Taking $a_i = X_i$, $b_i = 1$, $y = g/2$ and $z = (g/2) - \alpha$ it follows from Lemma 2.2 that

$$\sum_{i=1}^{N} X_i^{g-\alpha} \sum_{i=1}^{N} X_i^\alpha$$

increases with $|(g/2) - \alpha|$ for any real g (in most of the situations met in practice, g is found to be between 1 and 2). We now have

Theorem 2.1. The strategy consisting of the πPMSα

design and the corresponding Horvitz-Thompson estimator is inferior in the Θ_g-sense, to the strategy consisting of the πPS design and the Horvitz-Thompson estimator corresponding to this design whenever α does not lie between $g-1$ and 1.

Proof. Follows from Lemmas 2.1 and 2.2 and the fact that

$$\sum_{i=1}^{N} X_i^{g-\alpha} \sum_{i=1}^{N} X_i^{\alpha} > \sum_{i=1}^{N} X_i^{g-1} \sum_{i=1}^{N} X_i$$

when $|(g/2)-\alpha| > |(g/2)-1|$.

Remark 2.1. For a given $g=g_0>1$ it can be seen that πPX strategy is better than πPX^{g_0} and πPX^{g_0-1} strategies. Nothing can however be said about $\pi PX^{g_0/2}$, since $f(g_0/2)<0$.

Remark 2.2. Rewriting (2.7) as $(\mathcal{E}_{\pi PMS_\alpha}-\mathcal{E}_{\pi PS})/\sigma^2 = (f(\alpha)/\nu_0)+\Delta(\alpha)$, it is however not difficult to obtain the condition under which the modified measure of size would be better.

3. A NEW CLASS OF DESIGNS

It is seen in the preceding discussion that since $f(g/2)<0$, it is not known if the $\pi PMS\alpha$ sampling schemes with $\alpha=g/2$ would fare better than the πPS schemes. But, at the same time we notice that such a $\pi PMS(g/2)$ scheme enables us to minimize the first term of (2.2) and this then motivates the search for a class of designs which are best suited for the use of $\hat{Y}_{HT}$ as an estimator of the population total, under the assumptions of (1.1).

Given the expected cost, ν_0 is fixed, we search for an optimum amongst the class of designs for which

$$\pi_i = cX_i^{g/2}, \quad i=1,2,\cdots,N \tag{3.1}$$

where c is given by $c=\nu_0/\sum_{i=1}^{N} X_i^{g/2}$. Under the criteria of unbiasedness and minimum variance $\mathrm{Var}(\hat{Y}_{HT})$ can not be uniformly minimized w.r.t. Y_i's. Now, following Hanurav [8] we have

$$\mathcal{E}_{\theta(g)}\ \mathrm{Var}(\hat{Y}_{HT}) = \sigma^2\sum_{i=1}^{N}\frac{\pi_i(1-\pi_i)}{c^2} + \frac{a^2}{c^2}\mathrm{Var}\{\textstyle\sum_{i\in s} X_i^{1-(g/2)}\} \tag{3.2}$$

and this implies that (3.2) would be a minimum when $\sum_{i\in s} X_i^{1-(g/2)}$ is a constant. (Working out on the lines of Hanurav [8] one would first show that minimization of (3.2) corresponds to minimization of

$$\sum_{i\neq j}^{N}\sum^{N} \pi_{ij}(\pi_i\pi_j)^{(1/g)-1}$$

which again corresponds to the minimization of $\mathrm{Var}\{\sum_{i\in s} \pi_i^{(1/g)-1}\}$ implying thereby that $\sum_{i\in s} X_i^{1-(g/2)}$ be a constant.) At this stage it may be pointed out that when $g=2$, this condition reduces to $\nu(s)=$ a constant as obtained by Hanurav.

Thus we have established the following:

Theorem 3.1. Let D be the class of designs with $\pi_i \propto X_i^{g/2}$ in conjunction with which the Horvitz-Thompson estimator $\hat{Y}_{HT}$ is used for the estimation of the population total Y. In class D, the θ_g-optimum designs for any $\theta_g\in\Theta_g$ which are best suited for the use of the Horvitz-Thompson estimator are those that satisfy

$$\sum_{i\in s} X_i^{1-(g/2)} = \text{constant, } K \text{ say.}$$

Remark 3.1. The result stated in Theorem 3.1 leaves open the problem of construction of sampling designs (for brevity called as πPMS designs henceforth, the modified size being $Z_i'=X_i^{g/2}$, in the subsequent discussion) such that for a given ν_0,

(a) $\sum_{i\in s} Z_i = \sum_{i\in s} X_i^{1-(g/2)} = \nu_0\sum_{i=1}^{N} X_i/\sum_{i=1}^{N} X_i^{g/2}$ (since $K=\nu_0\sum_{i=1}^{N} X_i/\sum_{i=1}^{N} X_i^{g/2}$) and

(b) $\pi_i \propto X_i^{g/2}$.

As a starting point a scheme similar to Hanurav's [9] for $\nu_0=2$ may be suggested with P_i's replaced by P_i''s equal to $Z_i'/\sum_{i=1}^{N} Z_i'$, $i=1,2,\cdots,N$ and the stopping rule being $\sum_{i\in s} Z_i=K$, but its properties are to be further investigated.

Remark 3.2. We illustrate now by providing an example that the class of πPMS designs is non-empty. Let $g=1.5$ and consider a population consisting of four units with auxiliary information X_i, $i=1,\cdots,4$. Let ν_0 be the

AN ILLUSTRATIVE EXAMPLE

U_i	X_i	$X_i^{g/2}$	$Z_i = X_i^{1-(g/2)}$
U_1	1	1	1
U_2	16	8	2
U_3	1	1	1
U_4	16	8	2
Total	34	18	6

expected cost fixed beforehand $=36/17$. Then we have $K=4$. We now need to construct a design for which $\sum_{i\in s} Z_i=4$ and $\pi_i\propto X_i^{g/2}$. We also impose the restriction that the variance of the estimate be estimable. Consider the following design $D(S, P)$:

s	p_s
U_1, U_2, U_3	1/17
U_1, U_3, U_4	1/17
U_2, U_4	15/17
$\sum_s p_s$	1

for which we have $\sum_{i\in s} Z_i = K = 4$ as required and furthermore, $\pi_1=2/17=\pi_3$ and $\pi_2=16/17=\pi_4$ which are $\propto X_i^{g/2}$ and the variance is estimable since $\pi_{ij}>0$ for all i and j.

Remark 3.3. In conclusion we remark that the strategy consisting of πPMS design (of Theorem 3.1) and the corresponding Horvitz-Thompson estimator in addition to being superior to the strategy consisting of a πPS design (since $\Delta(g/2)=0$ in (2.6)) and the associated Horvitz-Thompson estimator has a further advantage that the estimator used is still the Horvitz-Thompson estimator and preserves all its optimum properties (see for example Godambe [4], Godambe and Joshi [5], Hanurav [9], [10], Rao, T. J. [12], [13] and Vijayan [15]). It is also shown elsewhere (Rao, T. J. [14]) that the πPMS strategy with the corresponding Horvitz-Thompson estimator is superior to the Symmetrised Des Raj strategy under a general super population set up for *all* values of the parameters g, thereby settling the controversy regarding these two estimators.

REFERENCES

[1] Callebaut, D. K., "Generalization of Cauchy-Schwartz Inequality," *Journal of Mathematical Analysis and Applications*, 12 (December 1965,) 491–4.

[2] Cochran, W. G., "Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations," *Annals of Mathematical Statistics*, 17 (June 1946), 164–77.

[3] Godambe, V. P., "A Unified Theory of Sampling from Finite Populations," *Journal of Royal Statistical Society*, Ser. B, 17, No. 2 (1955), 269–78.

[4] ——, "An Admissible Estimate for Any Sampling Design," *Sankhya*, Ser. A, 22 (June 1960), 285–88.

[5] —— and Joshi, V. M., "Admissibility and Bayes Estimation in Sampling Finite Populations—I," *Annals of Mathematical Statistics*, 36 (December 1965), 1707–22.

[6] Hansen, M. H. and Hurwitz, W. N., "On the Theory of Sampling from Finite Populations," *Annals of Mathematical Statistics*, 14 (December 1943), 333–62.

[7] —— and Hurwitz, W. N., "On the Determination of Optimum Probabilities in Sampling," *Annals of Mathematical Statistics*, 20 (September 1949), 426–32.

[8] Hanurav, T. V., "On Horvitz-Thompson Estimator," *Sankhyā*, Ser. A, 24 (November 1962), 429–36.

[9] ——, "Optimum Utilisation of Auxiliary Information-πPS Sampling of Two Units from a Stratum," *Journal of Royal Statistical Society*, Ser. B, 29, No. 2 (1967), 374–91.

[10] ——, "Hyper-Admissibiilty and Optimum Estimators for Sampling Finite Populations," *Annals of Mathematical Statistics*, 39 (April 1968), 621–42.

[11] Horvitz, D. G. and Thompson, D. J., "A Generalization of Sampling without Replacement from a Finite Universe," *Journal of American Statistical Association*, 47 (December 1952), 663–85.

[12] Rao, T. J., "On the Choice of a Strategy for Ratio Method of Estimation," *Journal of Royal Statistical Society*, Ser. B, 29, No. 2 (1967), 392–7.

[13] ——, "On the Allocation of Sample Size in Stratified Sampling," *Annals of Institute of Statistical Mathematics*, 20, No. 1 (1968), 159–66.

[14] ——, "Horvitz-Thompson and Desraj Estimators Revisited," Research Report No. 18, Statistical Laboratory, University of Manchester, 1970.

[15] Vijayan, K., "On Horvitz-Thompson and Desraj Estimators," *Sankhyā*, Ser. A, 28 (March 1966), 87–92.