ON THE ALLOCATION OF SAMPLE SIZE IN STRATIFIED SAMPLING

T. J. RAO

(Received March 3, 1967; revised Feb. 7, 1968)

Summary

The problem of optimum allocation of sample size to strata is examined in the light of *a priori* distributions. In this context, we discuss with illustrations the justification for the assumption that the unknown proportionate values of σ_i^2's can be replaced by the proportionate values of the known α_i^2's, which are estimates of σ_i^2's. The sample size is allocated so as to minimize the expected variance of the strategy consisting of πPS sampling scheme and the Horvitz-Thompson estimator under a general super-population model. It is further shown that, in the sense of expected variance, πPS sampling for unstratified sampling is inferior to πPS stratified sampling with this type of allocation, unless the super-population parameter g attains the value 2, in which case both schemes are equivalent.

1. Introduction

In the case of Neyman's optimum allocation ([12]) of sample size to strata, we have the allocation given by $n_{i,\text{opt.}}=nN_i\sigma_i/\sum_{i=1}^{h} N_i\sigma_i$, where N_i is the size of the ith stratum ($i=1, 2,\cdots, h$; n is the total sample size and σ_i^2 is the within-variance for the ith stratum). Computation of $n_{i,\text{opt.}}$'s requires at least the proportionate values of σ_i^2's, which are unknown. In practice, some estimates α_i^2's of σ_i^2's are substituted. These estimates, usually, are the σ_i^2's of some auxiliary information closely related to the characteristic under study. In most of the cases, the values of the same characteristic studied in a previous occasion are treated as the auxiliary information. The justification for the assumption that the unknown proportionate values of σ_i^2's are usually not quite different from the proportionate values of the known α_i^2's is examined in the light of an *a priori* distribution by Hanurav [5] and here it is further developed.

2. Optimum allocation of sample size and prior distributions

Let Δ_g be the class of all prior distributions ∂ for which we have

$$\text{i)}\quad \mathcal{E}_\partial(Y_i \mid X_i)=aX_i\ ,$$

$$\text{ii)}\quad \mathcal{V}_\partial(Y_i \mid X_i)=\sigma^2 X_i^g\ ,\qquad 1\leqq g\leqq 2\ ,$$

$$\text{iii)}\quad \operatorname{Cov}_\partial(Y_i,\ Y_j \mid X_i, X_j)=0\ .$$

Then,

$$\sigma_i^2=\frac{1}{N_i}\left[\sum_{j=1}^{N_i} Y_{ij}^2-\frac{Y_i^2}{N_i}\right]$$

where Y_{ij} is the value of the characteristic $\mathcal{Y}$ for the jth unit of the ith stratum ($j=1, 2, \cdots, N_i$; $i=1, 2, \cdots, h$). We then have

$$\begin{aligned}\mathcal{E}_\partial(\sigma_i^2)&=\frac{1}{N_i}\left[\sum_j (a^2X_{ij}^2+\sigma^2X_{ij}^g)-\frac{1}{N_i}(a^2X_i^2+\sigma^2\sum_j X_{ij}^g)\right]\\&=\frac{1}{N_i}\left[a^2\left(\sum_j X_{ij}^2-\frac{X_i^2}{N_i}\right)+\sigma^2\left(1-\frac{1}{N_i}\right)\sum_j X_{ij}^g\right]\\&=a^2\alpha_i^2+\sigma^2\frac{N_i-1}{N_i}\left(\sum_j X_{ij}^g/N_i\right)\\&\cong a^2\alpha_i^2+\sigma^2\sum_j X_{ij}^g/N_i\ ,\end{aligned}$$

assuming $\dfrac{N_i-1}{N_i}$'s to be nearly equal to 1,

$$\begin{aligned}&=a^2\alpha_i^2+\sigma^2\left[(\sum_j X_{ij}^g/N_i)-(\sum_j X_{ij}^2/N_i)+\alpha_i^2+\bar{X}_i^2\right]\\&=(a^2+\sigma^2)\alpha_i^2+\sigma^2\left[\bar{X}_i^2+\frac{1}{N_i}(\sum_j X_{ij}^g-\sum_j X_{ij}^2)\right]\ .\end{aligned}$$

Thus σ_i^2's can be expected to be in the same proportion as α_i^2's if α_i^2's are proportional to $\bar{X}_i^2+\dfrac{1}{N_i}(\sum_j X_{ij}^g-\sum_j X_{ij}^2)$ i.e., if the squares of the corrected coefficients of variation (c.c.v.) of $\mathcal{X}$-character, defined by $\dfrac{\alpha_i^2}{\bar{X}_i^2-\partial_i/N_i^2}$, where $\partial_i=N_i(\sum_j X_{ij}^2-\sum_j X_{ij}^g)$ are equal in all the strata and when this condition is satisfied the allocation is given by

$$n_i = n \frac{N_i \sqrt{\bar{X}_i^2 - \dfrac{\partial_i}{N_i^2}}}{\sum_{i=1}^{h} N_i \sqrt{\bar{X}^2 - \dfrac{\partial_i}{N_i^2}}}$$

$$= n \frac{\sqrt{T_i^2 - \partial_i}}{\sum_{i=1}^{h} \sqrt{T_i^2 - \partial_i}} , \qquad \text{where} \quad T_i = \sum_j X_{ij} .$$

Thus we have

THEOREM 1. *Under the super-population model Δ_g, Neyman's optimum allocation reduces to allocation proportional to $\sqrt{T_i^2 - \partial_i}$, where $T_i = \sum_{j=1}^{N_i} X_{ij}$ is the total of the auxiliary variate for the ith stratum and $\partial_i = N_i(\sum_{j=1}^{N_i} X_{ij}^2 - \sum_{j=1}^{N_i} X_{ij}^g)$, when the corrected coefficients of variation of the $\mathcal{X}$-character are equal in all the strata.*

From the above theorem follows immediately

THEOREM 2. (See [5].) *Neyman's optimum allocation reduces to allocation proportional to the stratum totals of the auxiliary variate $\mathcal{X}$, under Δ_2, when the coefficients of variation of $\mathcal{X}$-character are equal in all the strata.*

PROOF. Under Δ_2, we have $\partial_i = 0$,

hence,

$$n = nT_i / \sum_{i=1}^{h} T_i = nT_i / T \qquad \text{etc.}$$

3. Illustrations and remarks

Based on theorem 2, optimum allocation can be approximated by allocation proportional to the stratum totals under a particular model (Δ_2) when the condition that the within-strata coefficients of variation are more or less equal is satisfied. Especially, when stratification is based on geographical contiguity or administrative and operational considerations this condition is mostly satisfied. Mahalanobis [10] proposed equalization of the stratum totals along with equal allocation as an approximation to optimum allocation. Hansen, Hurwitz and Madow [7] have demonstrated the use of this method for slight adjustments of strata sizes. Kitagawa [9] analyzed further and justified the principle of equipartition. Considering the allocation of a sample of 1000 saw mills, Hansen, Hurwitz and Madow [7] show that allocation proportional to stratum totals serves as an approximation to optimum allocation. Murthy [11] considered live data on 128 villages of an Indian tehsil and Table 1 shows the results and allocation.

Table 1

No. of strata=2

Method of stratification	Relative variance for allocation proportional to	
	T_i	$N_i\sigma_i$
I	.005266	.005005
II	.005052	.004773
III	.004878	.004702

Next, in allocating a sample of 1000 farms in the province of Söedermanland, Sweden, Dalenius [2] considered various types of allocations for four acreage groups. Allocation proportional to total acreage (termed 'value-allocation') is found to offer a good compromise to minimum variance allocation. Hansen, Hurwitz and Madow [7] discuss this value-allocation and minimum variance allocations based on data for various crops. In fact it is interesting to observe that the minimum variance allocation for autumn wheat and the value-allocation are as close as (260, 278, 311, 151) and (268, 263, 313, 156).

In most of the situations one assumes the model Δ_2, but there are populations for which we have to consider a more general model Δ_g and for completeness and general academic interest, theorem 1 is proved which is an extension of the above case and when one has an idea of the value of g then an allocation as suggested therein would be more apt. We illustrate this considering data on the 1920 inhabitants of 64 large cities in the United States, grouped into 2 strata (Cochran [1]). Table 2 gives the results when $g=1.75$.

Table 2

Stratum	Stratum size N_i	Stratum totals $T_i=\sum X_{ij}$	$\sum_j X_{ij}^2$
1	16	8349	4,756,619
2	48	7941	1,474,871

i	α_i	∂_i	$\sqrt{T_i^2-\partial_i}$	Allocation proportional to			
				α_i	T_i	$\sqrt{T_i^2-\partial_i}$	c.c.v.
1	173.8	60,655,664	3008.4	$.5n$	$.51n$	$.47n$	.93
2	57.9	51,822,528	3352.1	$.5n$	$.49n$	$.53n$	.84

4. Stratified πPS sampling

Consider a population consisting of h strata with N_i as the ith stratum, $i=1, 2, \cdots, h$. Let Y_{ij} and X_{ij} be the values of the $\mathcal{Y}$ characteristic and the $\mathcal{X}$ characteristic (auxiliary information) respectively, for the jth unit of the ith stratum. Let a πPS (π_λ, the probability of inclusion of the λth unit, being proportional to size) sample (see Hanurav [6]) of size n_i be taken from the ith stratum, such that $\sum_{i=1}^{h} n_i=n$. Let π_{ij} be the probability of inclusion of the jth unit in the ith stratum in the sample. As an estimator of the population total consider the Horvitz-Thompson estimator (Horvitz and Thompson [8])

$$\hat{Y}=\sum_{i=1}^{h}\sum_{j=1}^{n_i} y_{ij}/\pi_{ij} . \tag{1}$$

THEOREM 3. *Under the super-population model Δ_g, allocation of the sample size to the strata which minimizes the expected variance of* (1) *is given by*

$$n_i=\frac{n\sqrt{T_i\sum_{j=1}^{N_i} X_{ij}^{g-1}}}{\sum_{i=1}^{h}\sqrt{T_i\sum_{j=1}^{N_i} X_{ij}^{g-1}}}$$

where T_i is the total of $\mathcal{X}$-values for the ith stratum.

PROOF. We have

$$\mathcal{E}_{\Delta_g} V(\hat{Y})=\sum_{i=1}^{h}\sum_{j=1}^{N_i}\left(\frac{1}{\pi_{ij}}-1\right)\sigma_{ij}^2 , \tag{2}$$

where $\sigma_{ij}^2=\sigma^2 X_{ij}^g$ (see Godambe [3], Hanurav [4]). We now, minimize (2) subject to the condition $\sum_{i=1}^{h} n_i=n$. Introducing the Lagrange multiplier λ, consider

$$\mathcal{E}_{\Delta_g} V(\hat{Y})+\lambda\left(\sum_{i=1}^{h} n_i-n\right)=\sum_{i=1}^{h}\sum_{j=1}^{N_i}\left(\frac{1}{n_i P_{ij}}-1\right)\sigma_{ij}^2+\lambda\left(\sum_{i=1}^{h} n_i-n\right) ,$$

where

$$P_{ij}=X_{ij}/T_i .$$

Differentiating with respect to n_i and equating to zero, we have

$$-\frac{1}{n_i^2}\sum_{j=1}^{N_i}\frac{\sigma_{ij}^2}{P_{ij}}+\lambda=0 ,$$

which gives

$$n_i = \frac{1}{\sqrt{\lambda}} \sqrt{\sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{P_{ij}}}$$

and thus

$$n_i = n\sqrt{\sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{P_{ij}}} \Big/ \sum_{i=1}^{h} \sqrt{\sum_{j=1}^{N_i} \frac{\sigma_{ij}^2}{P_{ij}}}$$
$$= n\sqrt{T_i \sum_{j=1}^{N_i} X_{ij}^{g-1}} \Big/ \sum_{i=1}^{h} \sqrt{T_i \sum_{j=1}^{N_i} X_{ij}^{g-1}} \ ,$$

hence, we have the theorem.

DEFINITION 1. The allocation of sample size to the strata given above is called Δ_g-optimum allocation.

THEOREM 4. *Under the super-population model Δ_2, allocation of the sample size to the strata proportional to the stratum totals of the $\mathcal{X}$-variate minimizes the expected variance of* (1).

PROOF. The proof follows directly from theorem 3, by putting $g=2$.

Let n_i be the Δ_2-optimum sample size for the ith stratum and select the sample by a πPS sampling scheme within each stratum. Then as an estimator of the population total consider the estimator (1)

$$\hat{Y} = \sum_{i=1}^{h} \sum_{j=1}^{n_i} \frac{y_{ij}}{\pi_{ij}} = \sum_{i=1}^{h} \sum_{j=1}^{n_i} \frac{y_{ij}}{(n_i x_{ij}/T_i)} \ .$$

We now have the following theorem.

THEOREM 5. *In the sense of expected variance, under Δ_g, πPS unstratified sampling is inferior to πPS stratified sampling with Δ_g-optimum allocation and for $g=2$, both the schemes are equivalent.*

PROOF. We have

$$\mathcal{E}_\delta V(\hat{Y}_{\text{st.}}) = \sum_{i=1}^{h} \sum_{j=1}^{N_i} [\pi_{ij}^{-1} - 1]\sigma^2 X_{ij}^g \ ,$$

where

$$\pi_{ij} = \frac{n\mu_i X_{ij}}{X_i} \qquad \text{and} \qquad \mu_i = \frac{\sqrt{X_i \sum_j X_{ij}^{g-1}}}{\sum_{i=1}^{h} \sqrt{X_i \sum_j X_{ij}^{g-1}}} \ ,$$

and

$$V(\hat{Y}_{\text{unst.}})=\sum_{i=1}^{h}\sum_{j=1}^{N_i}[\pi_{ij}'^{-1}-1]\sigma^2 X_{ij}^g ,$$

where

$$\pi'_{ij}=nX_{ij}/X .$$

Thus,

$$\begin{aligned}\mathcal{E}_s\left[\frac{V(\hat{Y}_{\text{unst.}})-V(\hat{Y}_{\text{st.}})}{\sigma^2}\right]&=\sum_{i=1}^{h}\sum_{j=1}^{N_i}\left[\frac{X_{ij}^g}{\pi'_{ij}}-\frac{X_{ij}^g}{\pi_{ij}}\right]\\&=\sum_i\sum_j\left[\frac{X_{ij}^g}{(nX_{ij}/X)}-\frac{X_{ij}^g}{(n\mu_i X_{ij}/X_i)}\right]\\&=\frac{X}{n}\sum_i\sum_j X_{ij}^{g-1}-\sum_i\frac{X_i}{n\mu_i}\sum_j X_{ij}^{g-1}\\&=\frac{X}{n}\sum_i\sum_j X_{ij}^{g-1}-\left(\sum_i\sqrt{X_i\sum_j X_{ij}^{g-1}}\right)^2\Big/n\\&=\frac{1}{n}\left[\sum_{i=1}^{h}X_i\sum_{i=1}^{h}\left(\sum_{j=1}^{N_i}X_{ij}^{g-1}\right)-\left(\sum_{i=1}^{h}\sqrt{X_i}\sqrt{\sum_{j=1}^{N_i}X_{ij}^{g-1}}\right)^2\right]\\&\geqq 0 ,\end{aligned}$$

by Cauchy-Shwartz's inequality.

It is easily seen that when $g=2$, equality occurs. Hence we have the theorem.

5. Acknowledgement

The author is grateful to the referee for his critical comments on an earlier version of this paper.

INDIAN STATISTICAL INSTITUTE, CALCUTTA

REFERENCES

[1] W. G. Cochran, *Sampling Techniques*, John Wiley, New York, 1953.
[2] T. Dalenius, "Anvaendmingen av Stickprovsmetodik vid 1950 Ars Arealinventering och Kreatursraeknig," *Jordbruksekonomiska Meddelanden*, (1951), 59-75.
[3] V. P. Godambe, "A unified theory of sampling from finite populations," *Jour. Roy. Stat. Soc.*, 17B (1955), 269-278.
[4] T. V. Hanurav, "On Horvitz and Thompson's estimator," *Sankhyā*, A 24 (1962), 429-436.
[5] T. V. Hanurav, "*Optimum Sampling Strategies and Related Problems*," Ph. D. thesis, Indian Statistical Institute, 1965.
[6] T. V. Hanurav, "Optimum utilization of auxiliary information: πPS sampling of two units from a stratum," *J. R. Statist. Soc.*, 29B (1967), 374-391.

[7] M. H. Hansen, W. N. Hurwitz and W. G. Madow, *Sample Survey Methods and Theory*, John Wiley, New York, 1953.

[8] D. G. Horvitz and D. J. Thompson, "A generalization of sampling without replacement from a finite universe," *J. Amer. Statist. Ass.*, 47 (1952), 663-685.

[9] T. Kitagawa, "Some contributions to the design of sample surveys," *Sankhyā*, 17 (1956), 1-36.

[10] P. C. Mahalanobis, "Some aspects of the design of sample surveys," *Sankhyā*, 12 (1952), 1-7.

[11] M. N. Murthy, *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, 1967.

[12] J. Neyman, "On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection," *J. R. Statist. Soc.* 97 (1934), 558-606.