

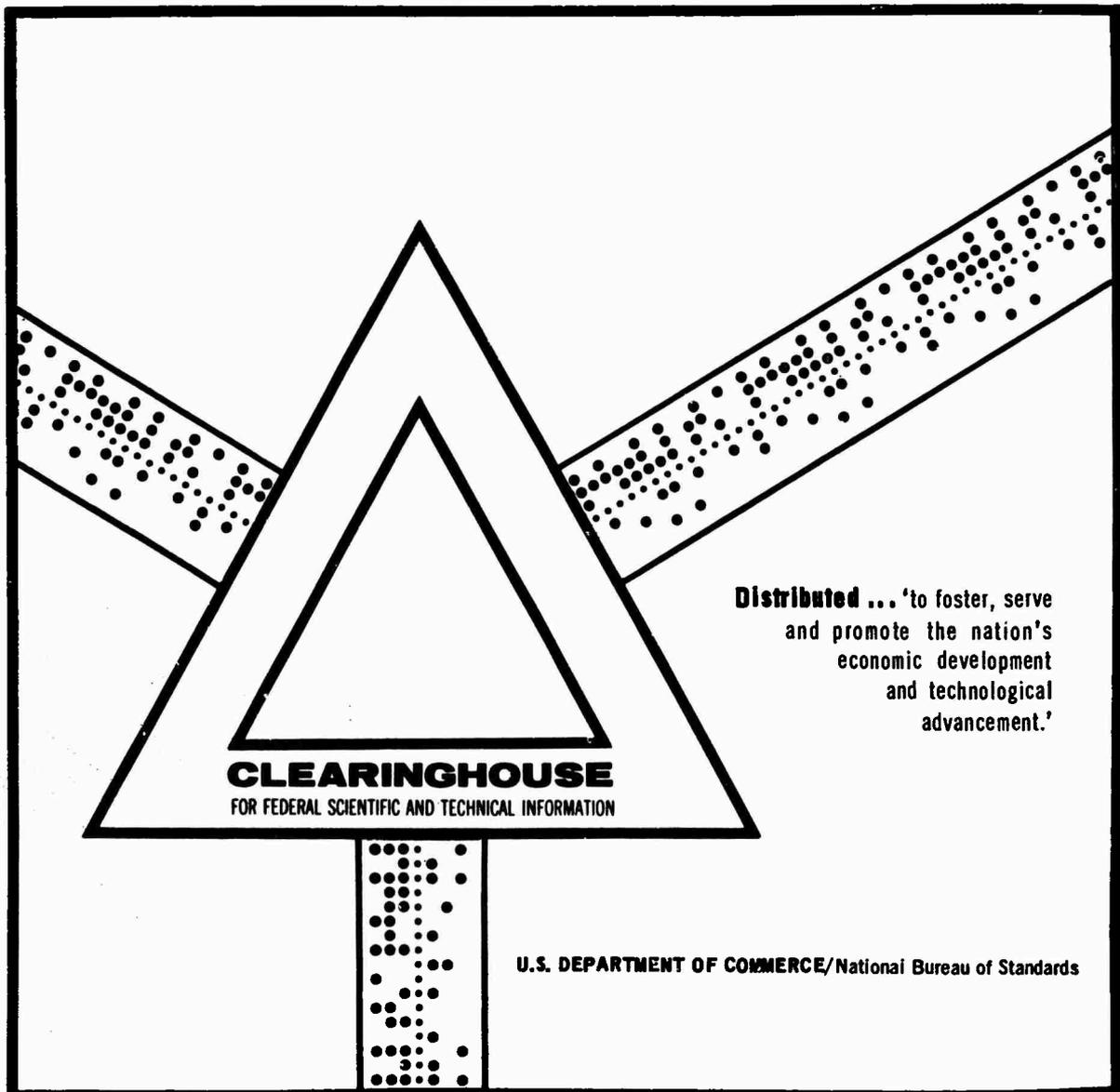
AD 698 484

A NEW ESTIMATION THEORY FOR SAMPLE SURVEYS, II

H. O. Hartley, et al

Texas A and M University  
College Station, Texas

1969



This document has been approved for public release and sale.

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)  Texas A & M University	2. REPORT SECURITY CLASSIFICATION  Unclassified	
	2b. GROUP  NA	

3. REPORT TITLE  
  
A New Estimation Theory for Sample Surveys, II

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)  
Reprint

5. AUTHOR(S) (First name, middle initial, last name)  
H. O. Hartley  
J. N. K. Rao

6. REPORT DATE 1969	7a. TOTAL NO. OF PAGES 19	7b. NO. OF REFS 7
------------------------	------------------------------	----------------------

8. CONTRACT OR GRANT NO. DAHCO4 69 C 0058	9a. ORIGINATOR'S REPORT NUMBER(S)  NA
9. PROJECT NO. c. 20061102B14C d.	
9b. OTHER REPORT NO(S) (Any other numbers that may be assigned to this report) 8049.4-M	

10. DISTRIBUTION STATEMENT  
  
Distribution of this report is unlimited.

11. SUPPLEMENTARY NOTES  None	12. SPONSORING MILITARY ACTIVITY U.S. Army Research Office-Durham Box CM, Duke Station Durham, North Carolina 27706
-------------------------------------	--

13. ABSTRACT

(AD 318 466)

This paper is a sequel to an earlier one (Hartley and Rao, 1968) on the same topic. Our first paper was predominantly concerned with simple random sampling (with or without replacement) from a finite population. In the present paper we are concerned with examining the relation of our findings to the more complex sampling procedures such as unequal probability sampling as well as stratified and multistage sampling.

↑

This document is approved for public release and wider distribution is unlimited.

14. KEY WORDS  
Estimating  
Sampling  
Statistical samples  
Population (statistics)

Reproduced by the  
CLEARINGHOUSE  
for Federal Scientific & Technical  
Information Springfield Va. 22151

AD 698484

D D  
REG  
JAN 5  
REG

ARO-D PROPOSAL

No. CRDARD-M-8049

"New Developments in Sample Survey Theory"

DAHCO4 69 C 0058

TECHNICAL REPORT NO. 2

"A New Estimation Theory for Sample Surveys, II"

by H. O. Hartley and J.N.K. Rao

## A New Estimation Theory for Sample Surveys, II

H. O. Hartley and J. N. K. Rao

Texas A & M University

### 1. Introduction.

This paper is a sequel to an earlier one (Hartley and Rao, 1968) on the same topic. Accordingly, it will be necessary to briefly recall the basic results of the earlier paper and relate that paper to the present one. Our first paper was predominantly concerned with simple random sampling (with or without replacement) from a finite population. In the present paper we are concerned with examining the relation of our findings to the more complex sampling procedures such as unequal probability sampling as well as stratified and multi-stage sampling.

The basic feature of our theory was a special parametrisation of a finite population of  $N$  units with  $k$  characteristics attached to each unit. Denote by  $\underline{y}_i$  the  $k$ -vector attached to the  $i$ -th unit. We assume that all elements of the  $\underline{y}_i$  are measured on discrete scales and that only a finite set of  $T$  measurement vectors  $\underline{y}_t$  ( $t = 1, 2, \dots, T$ ) are possible for the  $\underline{y}_i$ . Denote then by

$$N_t = \text{no. of units in the population having } \underline{y}_t \quad (1)$$

satisfying the conditions

$$N_t \geq 0 \text{ and } \sum_{t=1}^T N_t = N. \quad (2)$$

Henceforth, sums and products for  $t$  are over  $1, 2, \dots, T$ .

The parameters  $N_1, \dots, N_T$  completely describe any finite population. The number  $T$  is usually large although sometimes occasions arise when  $T$  is small

or moderate and the estimation of the  $N_t$  is of intrinsic interest, as for example when the  $N_t$  represent a frequency distribution such as the number of households in the community comprising  $t$  persons. However, in most cases we shall be concerned with the estimation of a few simple parametric functions such as the population moments and not with the separate estimation of the excessively large number of parameters  $N_t$ .

Finite population sampling will normally consist of (a) the sample design, i.e., the procedure of drawing a sample of  $n$  distinct units (where  $n$  may be fixed or random) and with measuring the  $y_t$  for these units, (b) the use of the measured  $y_t$  to compute estimators of the population parameters.

In our previous paper we restricted (a) to simple random sampling and we confined the computation of estimators (b) to what we termed 'scale-load' estimators. These were defined as mathematical functions of the scale vectors  $y_t$  and of their sample loads (frequencies)  $n_t = \text{no. of units in the sample having } y_t$ . Thus any identifying labels,  $i$ , that may be attached to the units may or may not be used for the implementation of the sample design; however, labels are not directly used in the computation of the estimators. Nevertheless, in situations where the labels,  $i$ , are observable characteristics of the units and are considered informative observables, the labels may be adjoined to the vectors  $y_i$  as a  $(k + 1)$ -th component.

We were able to show that within the class of 'scale-load' estimators many of the estimators in current use possess interesting optimality properties in simple random sampling. Specifically the estimators are either UMV (unbiased minimum variance) or maximum likelihood estimators or both. Some of these results are briefly restated in Section 2. In the remaining sections of the present paper we are

concerned with the role these results play in the more complex sampling procedures. Briefly our findings are: (1) The above parametrization of finite populations will continue to yield useful likelihood formulations for sampling designs providing maximum likelihood and Bayesian estimation procedures. UMV property will be the exception rather than the rule. (2) We consider that identifying labels of primary units (or all but the last stage units) will often be available as well as informative. There are, however, situations in which higher stage units are not labelled as is the case, for example, for certain subsets of machine parts produced in bulk, the water supply of water works produced during certain time periods, etc. Certain situations where labels of higher stage units are not informative also exist, for example identifiable subsets of certain lists. Both 'scale-load' and label-dependent estimators are therefore required. As would be expected, there is usually no UMV estimator in the class of label-dependent estimators. (3) A particular problem arises when label dependence of estimators is used in conjunction with Bayesian concepts and separate prior distributions are allowed for the individually identifiable units. The resulting posterior distributions and hence Bayesian inferences do not depend on the survey design which in the frame work of Bayesian theory becomes a randomization procedure irrelevant in making posterior inferences. However, the absurd result that Bayesian theory leads to when applied to simple sampling or ultimate-stage unit sampling (Godambe, 1966) is perhaps our strongest point in favor of examining estimators that do not depend on the labels of the ultimate-stage units.

## 2. Simple random sampling.

If a simple random sample of fixed size  $n$  is drawn without replacement

from the population of  $N$  units, the likelihood of the  $n_t$  is given by

$$L(N_1, \dots, N_T) = \prod \binom{N}{n_t} / \binom{N}{n} \quad (3)$$

where  $n_t \geq 0$  and  $\sum n_t = n$ . We confine ourselves here to the case of a single character  $y$  attached to the units (i.e.,  $k = 1$ ). In our previous paper we have shown that any function of the  $n_t$  is an UMV estimator of its expectation.

Specifically some of the more important parametric functions and their UMV estimators are given below:

<u>Parametric function</u>	<u>UMV estimator</u>
$N_t/N$	$n_t/n$
$\mu'_r = N^{-1} \sum N_t y_t^r$	$m'_r = n^{-1} \sum n_t y_t^r$ <span style="float: right;">(4)</span>
$\sigma^2 = \mu'_2 - \mu_1'^2$	$\frac{n(N-1)}{N(n-1)} (m_2^i - m_1'^2)$

Notice that the estimators do not depend on  $T$  or the non-observed  $y_t$ . When  $N/n$  is an integer,  $n_t/n$  and  $m'_r$  are also the maximum likelihood estimators (see the Appendix). When  $N/n$  is not integral, the maximization of (3) over the integral grid  $N_t$  can be achieved by the algorithm given in the Appendix; however, since UMV estimators exist, the maximum likelihood estimators may not have particular merit for small samples. The possibility of using maximum likelihood estimators of the  $N_t$  when  $T$  is small and the  $N_t$  are parameters of interest is being examined by a Monte Carlo study.

Turning now to Bayesian estimation, we have used in our previous paper the mathematically convenient prior distribution suggested by Hoadley (1968) and given by

$$\varphi(N_1, \dots, N_T) \propto \prod \frac{(N_t + v_t - 1)!}{N_t! (v_t - 1)!}, \quad v_t > 0. \quad (5)$$

The 'Bayes estimator' of  $\mu_r'$  is the posterior expectation of  $\mu_r'$  and is given by

$$E'(\mu_r') = (1 - \frac{n}{N})[w m_r' + (1 - w)M_r'] + \frac{n}{N} m_r' \quad (6)$$

where

$$w = n/(n + v), \quad v = \sum v_t \quad (7)$$

and

$$M_r' = v^{-1} \sum v_t y_t^r \quad (8)$$

It should be noted that the estimator (6) only requires the knowledge of  $M_r'$  (the prior mean of  $\mu_r'$ ) and  $w$ , i.e., in the case of  $r = 1$  the knowledge only of the prior mean  $M_1'$  and the relative weight  $w$  of the sample and prior information. Moreover, although the  $v_t$  are akin to a prior sample frequencies, the posterior mean is not simply the mean of the pooled 'sample'  $v_t + n_t$ . It duly recognizes the fact that, as  $n \rightarrow N$ , the sample mean  $m_1'$  will tend to  $\mu_1'$  and that the prior is ignored.

The expected loss which the decision maker faces by choosing the 'Bayes estimator' is given by the posterior variance

$$V'(\mu_r') = \frac{(N-n)(N+v)}{N^2(n+v+1)} \left[ w m_{2r}' + (1-w)M_{2r}' - \left\{ w m_r' + (1-w)M_r' \right\}^2 \right] \quad (9)$$

The 'Bayes estimator' of  $\sigma^2$  is given by

$$E'(\sigma^2) = \frac{(n+v-v/N)}{(1+v/N)} V'(\mu_1') + \frac{n}{N} \left[ m_2' - m_1'^2 + (1-\frac{n}{N})(1-w)^2 (m_1' - M_1')^2 \right] \quad (10)$$

It should be noted that, if the prior information is solely based on a pilot sample,  $M_r'$  and  $v$  would roughly represent the  $r$ -th sample moment based on the pilot sample and the pilot sample size respectively.

Turning to simple random sampling with replacement, suppose a random sample of fixed size  $m$  is drawn with equal probability and with replacement. Let

$n$  denote the number of distinct units in the sample and  $n_t$  the number of distinct units having the value  $y_t$  in the sample. The total likelihood is given by

$$L(N_1, \dots, N_T) = P(n) \frac{\prod \binom{N_t}{n_t}}{\binom{N}{n}} \quad (11)$$

where the probability  $P(n)$  is a function only of  $m$  and  $N$ . For this sample design no UMV exists, but the maximum likelihood estimator of  $\mu_r'$  is  $m_r' = n^{-1} \sum n_t y_t^r$  provided  $N = cn$  least common multiple of  $1, 2, \dots, m$  ( $c = \text{integer}$ ). In particular, the maximum likelihood estimator of the population mean  $\mu_1'$  is the sample mean based only on the distinct units in the sample and it is uniformly more efficient than the customary sample mean based on all the sample draws. With the prior distribution (5), the 'Bayes estimator' of  $\mu_r'$ , the posterior variance of  $\mu_r'$  and the 'Bayes estimator' of  $\sigma^2$  are respectively given by (6), (9) and (10), where  $n$  and the  $n_t$  are as defined above.

### 3. Estimation with concomitant variables.

In our earlier paper we have considered a situation customarily dealt with by ratio or regression method of estimation in which two variates  $y$  and  $x$  are attached to each of the units and the population mean  $\bar{Y}$  of 'target variate'  $y$  is to be estimated utilizing the available information about  $x$ . Assuming that only the population  $\bar{X}$  of  $x$  is known, we have shown that an approximation to the maximum likelihood estimator of  $\bar{Y}$  is closely related to the customary regression estimator, provided the sample size  $n$  is moderately large. In this section we extend this result to multiple concomitant variables  $x_1, \dots, x_k$ , assuming that only the population means  $\bar{X}_1, \dots, \bar{X}_k$  are known. We show that, for moderately large  $n$ , an approximation to the maximum likelihood estimator of  $\bar{Y}$  is closely

related to the customary multiple regression estimator.

As before, we assume that a finite set of  $T$  distinct, known values  $y_t$  are feasible for  $y$ . Likewise, we assume that  $I_j$  distinct, known values  $x_{ji}$  are feasible for  $x_j$  ( $j = 1, \dots, k$ ). Let  $N_{i_1 \dots i_k t}$  denote the number of units in the population which have  $x_{1i_1}, \dots, x_{ki_k}$  and  $y_t$  attached to them. Let  $n_{i_1 \dots i_k t}$  be the number of units in the simple random sample of size  $n$  (drawn without replacement) which have  $x_{1i_1}, \dots, x_{ki_k}$  and  $y_t$  attached to them.

We consider only the multinomial situation in which  $N \rightarrow \infty$  and  $N_{i_1 \dots i_k t} / N \rightarrow P_{i_1 \dots i_k t}$  while  $n$  is held fixed. The likelihood  $L$  is then given by the multinomial distribution with probabilities  $P_{i_1 \dots i_k t}$ . The restrictions on the  $P_{i_1 \dots i_k t}$  are given by

$$\underline{P} \geq 0, \underline{P}' \underline{i} = 1 \text{ and } \underline{P}' \underline{Z} = \underline{\bar{X}} \quad (12)$$

where  $\underline{P}'$  is the  $n \times 1$  vector of the  $P_{i_1 \dots i_k t}$ ,  $\underline{i}$  is the  $1 \times n$  vector of 1's,  $\underline{\bar{X}}' = (\bar{X}_1, \dots, \bar{X}_k)$  and  $\underline{Z} = (x_{1i_1}^* | \dots | x_{ki_k}^*)$  where  $\underline{P}' x_j^* = \bar{X}_j$  ( $j = 1, \dots, k$ ). As in our previous paper, it can be shown that for moderate sample sizes  $n$  the global maximum of the multinomial likelihood can only be attained if  $P_{i_1 \dots i_k t} = 0$  for all those variate combinations for which  $n_{i_1 \dots i_k t} = 0$ , and  $P_{i_1 \dots i_k t} > 0$  for the remainder. Confining then the maximization to the latter  $P_{i_1 \dots i_k t}$  only and introducing the Lagrangian multipliers  $\lambda$  and  $\underline{\mu}' = (\mu_1, \dots, \mu_k)$ , the maximization of  $\log L$  subject to (12) is attained for  $\underline{P} = \hat{\underline{P}}$  where

$$\hat{P}_{i_1 \dots i_k t} = \frac{n_{i_1 \dots i_k t}}{n} \left[ 1 + \frac{1}{n} \sum_{j=1}^k \mu_j (x_{ji} - \bar{X}_j) \right]^{-1}. \quad (13)$$

Expanding  $\hat{\underline{P}}' \underline{i} = 1$  to first three terms we obtain

$$n(\bar{\underline{x}} - \underline{\bar{X}})' \underline{\mu} = \underline{\mu}' \underline{\bar{X}}^* \underline{\bar{X}}^* \underline{\mu} \quad (14)$$

where  $\bar{\underline{x}}' = (\bar{x}_1, \dots, \bar{x}_k)$  is the vector of sample means and  $\underline{\bar{X}}^* \underline{\bar{X}}^* = \underline{S}^* = (s_{jp}^*)$

where

$$s_{jp}^* = n^{-1} \sum_{i_1 \dots i_k} n_{i_1 \dots i_k} (x_{ji_{i_1}} - \bar{x}_j)(x_{pi_{i_2}} - \bar{x}_p).$$

It is readily seen that the solution of (14) is given by

$$\underline{\mu} = n(\underline{\bar{X}}^* \underline{\bar{X}}^*)^{-1} (\bar{\underline{x}} - \underline{\bar{X}}). \quad (15)$$

Now using (15) and expanding (13) to the first two terms we get

$$\hat{\underline{P}} = \frac{1}{n} \left[ \underline{n} + \underline{X}' (\underline{\bar{X}}^* \underline{\bar{X}}^*)^{-1} (\bar{\underline{x}} - \underline{\bar{X}}) \right] \quad (16)$$

where  $\underline{n}$  is the  $1 \times n$  vector of the  $n_{i_1 \dots i_k}$  and  $\underline{X}'$  is given by  $\underline{y}' \underline{X}' = (s_{1y}^*, \dots, s_{ky}^*)$

where  $\underline{P}' \underline{y} = \bar{y}$  and

$$s_{jy}^* = n^{-1} \sum_{i_1 \dots i_k} n_{i_1 \dots i_k} y_t (x_{ji_{i_1}} - \bar{x}_j), \quad j = 1, \dots, k.$$

An improved approximation, along the lines of our previous paper, can be obtained by expanding (13) to the first three terms.

Using (16), an approximation to the maximum likelihood estimator of the population mean  $\bar{Y} = \underline{P}' \underline{y}$  is given by

$$\hat{\underline{Y}} = \hat{\underline{P}}' \underline{y} \doteq \bar{y} + (\bar{\underline{x}} - \underline{\bar{X}})' \underline{S}^{*-1} \underline{s}_{\cdot y}^* \quad (17)$$

where  $\underline{s}_{\cdot y}^*$  is the  $k$ -vector of the  $s_{jy}^*$ . The customary multiple regression estimator is given by

$$\hat{\underline{Y}}_r = \bar{y} + (\bar{\underline{x}} - \underline{\bar{X}})' \underline{S}^{-1} \underline{s}_{\cdot y} \quad (18)$$

where  $\underline{S} = (s_{jp})$ ,  $\underline{s}_{\cdot y}$  is the  $k$ -vector of the  $s_{jy}$  and

$$s_{jp} = s_{jp}^* - (\bar{x}_j - \bar{X}_j)(\bar{x}_p - \bar{X}_p)$$

$$s_{jy} = s_{jy}^* - \bar{y}(\bar{x}_j - \bar{X}_j).$$

Although (17) differs slightly from (18), the above development clearly shows that, at least in large samples, the customary multiple regression estimator is essentially the maximum likelihood estimator.

4. Stratified simple random sampling without replacement.

4.1 UMV Estimator.

Suppose there are  $L$  strata in the population with  $N_i$  units in the  $i^{\text{th}}$  stratum ( $i = 1, \dots, L$ ). Denote by  $N_{it}$  the number of units in the population belonging to the  $i^{\text{th}}$  stratum and having the measurement  $y_{it}$  ( $t = 1, \dots, T_i$ ) so that  $\sum_t N_{it} = N_i$  ( $\sum_i N_i = N$ ). A stratified simple random sample ( $n_1, \dots, n_L$ ) is drawn without replacement, ( $\sum_i n_i = n$ ), and  $n_{it}$  denotes the number of units in the sample belonging to the  $i^{\text{th}}$  stratum and having the measurement  $y_{it}$ , ( $\sum_t n_{it} = n_i$ ). Now the likelihood of the  $n_{it}$  is given by

$$L(N_{11}, \dots, N_{LT_L}) = \prod_{i=1}^L \left[ \frac{\binom{N_{i1}}{n_{i1}} \dots \binom{N_{iT_i}}{n_{iT_i}}}{\binom{N_i}{n_i}} \right]. \quad (19)$$

Therefore, the  $n_{it}$  are complete sufficient for the  $N_{it}$  and, hence, the UMV estimator of the population mean  $\bar{Y} = N^{-1} \sum_i \sum_t N_{it} y_{it}$  is the customary estimator

$$\hat{\bar{Y}} = N^{-1} \sum_{it} \hat{N}_{it} y_{it} = N^{-1} \sum_i N_i \bar{y}_i \quad (20)$$

where  $\hat{N}_{it} = (N_i/n_i) n_{it}$  is the UMV estimator of  $N_{it}$ . It also follows that the maximum likelihood estimators of the  $N_{it}$  and  $\bar{Y}$  are the UMV estimators  $\hat{N}_{it}$  and  $\hat{\bar{Y}}$  respectively, when the  $N_i/n_i$  are integral. Notice that each stratum is described by its separate set of parameters, i.e., we have an additional subscript  $i$  to index the strata.

An interesting special case occurs when the stratification is according to the size of the units, say  $x_i$ . If we assume that  $x_i$  is constant within strata and use the allocation proportional to total size, i.e.,

$$n_i = n(N_i x_i / \sum N_i x_i) = N_i P_i \quad (\text{say}) \quad \text{where } \sum P_i = n, \text{ we get}$$

$$\hat{Y} = N^{-1} \sum_{it} \hat{N}_{it} y_{it} = N^{-1} \sum_{it} \frac{n_{it} y_{it}}{P_i} \quad (21)$$

which is a 'Horvitz-Thompson' type estimator.

#### 4.2. Bayesian optimization of stratified sampling.

Ericson (1965) has presented a solution to the problem of optimum allocation when prior information in the form of a prior distribution is available.

He has, however, assumed: (a)  $N_i = \infty$ ,  $i = 1, \dots, L$ , (b) normality of the within stratum populations and (c) known within stratum population variances  $\sigma_i^2$ .

Assuming that the within stratum population means  $\mu_i$  have independent normal priors with means  $m_i$  and variances  $v_{ii}'$ , he has shown that the posterior variance of the population mean  $\mu = \sum \pi_i \mu_i$  is given by

$$v'' = \sum_i \left[ \pi_i^2 / \left( \frac{1}{v_{ii}'} + \frac{n_i}{\sigma_i^2} \right) \right] \quad (22)$$

where  $\pi_i$  is the known proportion of the population units falling within the  $i^{\text{th}}$  stratum. Ericson has given a computational algorithm to find  $n_i \geq 0$  ( $i = 1, \dots, L$ ) such that (22) is minimized subject to the cost constraint

$$\sum c_i n_i = C \quad (23)$$

where  $C$  is the given budget.

Recently, Draper and Guttman (1968) have relaxed the assumption (c) and presented a sequential allocation scheme which <sup>appears</sup> simpler than Ericson's al-

gorithm. They have also considered the case of unknown proportions  $\pi_i$ . Using our present approach, one of us (J. N. K. Rao, 1968) has given a solution which is free from the restrictive assumptions (b) and (c). Extension to multiple priors and/or multiple characteristics by the use of convex programming was also considered. In this section we present a complete solution by relaxing the assumption (a) also.

We assume that prior information on the  $N_{it}$  is available in the form of (5) for each  $i$  and that the priors are independent. Therefore, the prior distribution of  $N_{11}, \dots, N_{LT_L}$  is

$$\varphi(N_{11}, \dots, N_{LT_L}) \propto \prod_i \prod_t \frac{(N_{it} + v_{it} - 1)!}{N_{it}!(v_{it} - 1)!} \quad (24)$$

$$v_{it} > 0, \sum_t v_{it} = v_i$$

Now, since  $\bar{Y} = N^{-1} \sum_i N_i \bar{Y}_i$  where  $\bar{Y}_i$  is the  $i^{\text{th}}$  stratum population mean, we get using (6) and (9) the posterior mean of  $\bar{Y}$  as

$$E'(\bar{Y}) = N^{-1} \sum_i N_i \left[ \left(1 - \frac{n_i}{N_i}\right) \sum_t \frac{v_{it} + n_{it}}{n_i + v_i} y_{it} + \frac{n_i}{N_i} \sum_t \frac{v_{it}}{v_i} y_{it} \right] \quad (25)$$

and the posterior variance of  $\bar{Y}$  as

$$V'(\bar{Y}) = N^{-2} \sum_i N_i^2 \left(1 - \frac{n_i}{N_i}\right) \left(1 + \frac{v_i}{N_i}\right) (n_i + v_i + 1)^{-1} \cdot \left[ \sum_t \frac{n_{it} + v_{it}}{n_i + v_i} y_{it}^2 - \left( \sum_t \frac{n_{it} + v_{it}}{n_i + v_i} y_{it} \right)^2 \right] \quad (26)$$

Since the posterior variance (26) depends on the to be observed sample values  $n_{it}$ , we take the expectation of (26) with respect to the marginal distribution of the  $n_{it}$ . It follows from Hoadley (1968) that the marginal distribution of the  $n_{it}$  is given by

$$f(n_{11}, \dots, n_{LT_L}) = \prod_i \left[ \frac{\prod_t \binom{n_{it} + v_{it} - 1}{n_{it}}}{\binom{n_i + v_i - 1}{n_i}} \right] \quad (27)$$

which is identical to that in the case of infinite populations with Dirichlet prior distributions. Therefore, using the results of J. N. K. Rao (1968) it follows from (26) that the expected posterior variance of  $\bar{Y}$  is

$$V''(\bar{Y}) = \sum_i \frac{N_i^2}{N^2} \left(1 - \frac{n_i}{N_i}\right) \left(1 + \frac{v_i}{N_i}\right) \frac{A_i}{n_i + v_i} \quad (28)$$

where

$$\frac{A_i}{v_i} = (v_i + 1)^{-1} \sum_t \frac{v_{it}}{v_i} \left[ y_{it} - \left( \sum_t \frac{v_{it}}{v_i} y_{it} \right) \right]^2. \quad (29)$$

It follows, using (9) and (10), that

$$\text{Prior variance of } \bar{Y}_i = \left( \frac{1}{v_i} + \frac{1}{N_i} \right) A_i \quad (30)$$

and

$$\text{Prior mean of } S_i^2 = A_i \quad (31)$$

where  $N_i \sigma_i^2 = (N_i - 1) S_i^2$ .

Now (28) is a separable convex function in the  $n_i$  and, therefore, the values  $n_i$  which minimize (28) subject to (23) and  $0 \leq n_i \leq N_i$  ( $i=1, \dots, L$ ) can be obtained by convex programming\*. It is also possible to develop a sequential allocation procedure analogous to that of Draper and Guttman (1968).

It is important to note that the knowledge of the complete priors is not essential for the optimum allocation — only that of the prior mean of  $\sigma_i^2$  and prior variance of  $\mu_i$  is needed. If the priors are solely based on pilot samples within each stratum, then  $[(v_i+1)/v_i]A_i$  and  $v_i$  would roughly represent the pilot-sample variance and the pilot sample size respectively.

The extension of the above results to multiple priors and/or multiple

\* In our original version we ignored the restriction  $n_i \leq N_i$  and Ericson has pointed this out.

characteristics follows along the lines of J. N. K. Rao (1968) and the optimum allocation is obtained by convex programming.

5. Single-stage unequal probability sampling.

In the preceding sections we have been mainly concerned with sampling procedures in which all the units had an equal chance of selection. The only exception is stratified sampling (Section 4.1) in which strata allocations  $n_i$  proportional to the products  $N_i x_i$  gave all the  $N_i$  units in the  $i^{\text{th}}$  size stratum an equal inclusion probability of  $P_i = n(N_i x_i / \sum N_i x_i)$  which was varied from stratum to stratum. While unequal probability sampling by 'size strata' may be satisfactory for many practical purposes, situations often arise in which we desire to vary the inclusion probability from unit to unit. However, this type of unequal probability sampling mainly arises in the selection of primary sampling units in multi-stage sampling which we discuss in Section 6. Here we confine ourselves to the (rare) situations where unequal probability sampling is used in 'single-stage' or 'ultimate-stage' sampling of units which are not necessarily identifiable in advance of sampling.

As an example of p.p.s. sampling of this kind, we may mention here the sampling of farm operators in Iowa counties proportional to the land acreages they operate. If a county map can be covered by a rectangle with dimensions  $Z$  by  $W$  and  $(z_i, w_i), i = 1, \dots, r (r \geq m)$  denote uniform variables with  $0 \leq z_i \leq Z$  and  $0 \leq w_i \leq W$ , co-ordinates  $(z_i, w_i)$  can be pinpointed on the map and the interviewer can be instructed to ascertain (in order of draw) the first  $m$  operators whose land acreages contain the pinpointed land marks. This results in p.p.s. sampling with replacement in which  $p_i = x_i / X$  ( $x_i = \text{land}$

acreage of  $i^{\text{th}}$  operator only known for sampled operators,  $X$  = total land acreage known in advance of sampling) where  $p_i$  denotes the probability of selection of  $i^{\text{th}}$  operator at a single draw. This well-known situation of 'multinomial sampling' is the only one discussed in this section. We show that it can be reparametrized in such a way that optimality properties can be formulated for certain estimators.

Let  $r_i = y_i/p_i$  and denote by  $r_t$  ( $t = 1, \dots, T$ ) the set of  $T$  discrete scale points feasible for the  $r_i$ . Let the score  $m_i$  denote the number of times  $i^{\text{th}}$  unit is included in the sample ( $i = 1, \dots, N; \sum m_i = m$ ). We now classify the  $r_i$  into the  $T$  groups and denote by

$$p_{it} = \begin{cases} p_i & \text{if for the } i^{\text{th}} \text{ unit } r_i = r_t \\ 0 & \text{otherwise} \end{cases}$$

$$y_{it} = \begin{cases} y_i & \text{if for the } i^{\text{th}} \text{ unit } r_i = r_t \\ 0 & \text{otherwise} \end{cases}$$

$$m_{it} = \begin{cases} m_i & \text{if for the } i^{\text{th}} \text{ unit } r_i = r_t \\ 0 & \text{otherwise} \end{cases}$$

The multinomial distribution of scoring  $m$  multinomial scores into  $N$  classes with probabilities  $p_i$  may then be written in the form

$$L(p_{11}, \dots, p_{NT}) = \frac{m!}{\prod_{i,t} m_{it}!} \prod_{i,t} p_{it}^{m_{it}} \quad (32)$$

and may be reparametrised as follows:

$$p_t = \sum_i p_{it}$$

$$y_{it} = \begin{cases} p_{it}/p_t & \text{if } p_t > 0 \\ 1 & \text{if } p_t = 0 \end{cases} \quad (33)$$

so that

$$\sum_i y_{it} = 1 \quad \text{if } p_t > 0. \quad (34)$$

Writing  $m_t = \sum_i m_{it}$ , we may factorize (32) as

$$L(p_{11}, \dots, p_{NT}) = \left[ \frac{m_t!}{\prod_i m_{it}!} \prod_t p_t^{m_t} \right] \left[ \frac{\prod_t m_t!}{\prod_{i,t} m_{it}!} \prod_{i,t} y_{it}^{m_{it}} \right]. \quad (35)$$

Equation (35) shows that the  $m_t$  are sufficient for the  $p_t$  since the latter are only involved in the marginal distribution of the  $m_t$  and not in the conditional distribution of the  $m_{it}$  given the  $m_t$ .

The maximum likelihood estimators of the  $p_t$  are given by the ratios  $m_t/m$  and, hence, the maximum likelihood estimator of the population total

$$Y = \sum_i y_i = \sum_{i,t} y_{it} = \sum_t \sum_i y_{it} = \sum_t r_t p_t \quad (36)$$

is given by

$$\hat{Y} = \sum_t r_t \frac{m_t}{m} = \frac{1}{m} \sum_t r_t \sum_i m_{it} = \frac{1}{m} \sum_i \frac{y_i m_i}{p_i} \quad (37)$$

which is the customary unbiased estimator of  $Y$  in p.p.s. sampling with replacement.

Finally it should be noted that (35) is the likelihood for the scores which do not necessarily represent counts of distinct units in the population. However, it is possible to obtain the likelihood of the number of distinct units in the sample with scale ratio  $r_t$  which we denote by  $n_t$ . The distribution of the  $m_t$  is given by

$$L(p_1, \dots, p_T) = \frac{m_t!}{\prod_i m_{it}!} \prod_t p_t^{m_t} \quad (38)$$

and the conditional distribution of the  $n_t$  given the  $m_t$  can be obtained in terms of the  $Y_{it}$  from formula (4.3) of Kullback (1937). Finally the likelihood of the  $n_t$  can be obtained by summation of the product (i.e., the joint distribution) over  $m_t = n_t$  to  $m$  subject to  $\sum m_t = m$ . We intend to examine this distribution in more detail elsewhere.

Although only one single method of unequal probability sampling is examined in this section and although the method examined is known not to be particularly efficient, the discussion clearly indicates the possibility of deriving concrete likelihoods for other unequal probability sampling methods with the help of our technique of parametrisation.

#### 6. Two-stage sampling.

In order to simplify the discussion we confine ourselves to two-stage sampling in which the primaries are selected with equal or unequal probabilities. Consider then a population consisting of  $L$  primary units  $i = 1, \dots, L$  of which  $l$  will be sampled and denote by  $N_i$  the number of secondary units in the  $i^{\text{th}}$  primary. Denote by  $N_{it}$  the number of units in the  $i^{\text{th}}$  primary which have the scale value  $y_{it}$  ( $t = 1, \dots, T_i$ ) so that  $\sum_t N_{it} = N_i$ . Let  $u_i = 1$  if the  $i^{\text{th}}$  primary is in the sample and zero otherwise. Denote by  $P(u_1, \dots, u_L)$  the joint distribution of the  $u_i$  corresponding to the primary sampling procedure adopted and let  $n_i$  denote the number of secondary units to be drawn from the  $i^{\text{th}}$  primary if it is in the sample. The  $n_i$  are all specified a priori for  $i = 1, \dots, L$ . In this paper we only consider equal probability sampling of secondaries without replacement.

If we denote by  $n_{it}$  the number of secondaries having scale value  $y_{it}$  in the  $i^{\text{th}}$  sampled primary, then the joint likelihood of the  $u_i$  and the

$n_{it}$  is given by

$$L(N_{11}, \dots, N_{LT_L}) = P(u_1 \dots u_L) \prod_i \left[ \frac{\Gamma \binom{N_{it}}{u_i n_{it}}}{\binom{N_i}{u_i n_i}} \right]. \quad (39)$$

6.1. Maximum likelihood estimation.

We confine ourselves here to the case of  $N_i = \infty$ ,  $i = 1, \dots, L$ . The likelihood (39) reduces to

$$L(p_{11}, \dots, p_{LT_L}) = P(u_1, \dots, u_L) \prod_i \left[ \frac{(n_i u_i)!}{\prod_t u_i^{n_{it}}} \prod_t p_{it}^{u_i n_{it}} \right]. \quad (40)$$

Maximisation of (40) subject to  $\sum_t p_{it} = 1$  for  $i = 1, \dots, L$  leads to

$$\hat{p}_{st} = n_{st}/n_s \quad (\text{primary } s \text{ in the sample}) \quad (41)$$

while any values of  $p_{jt}$  are permissible for  $j$  not in the sample. The maximum likelihood solution will, therefore, in general not be unique. Furthermore, we do not have complete sufficiency here and, hence, no UMV estimator exists. We have not considered here 'scale-load' estimators which do not depend on primary labels.

6.2. Bayesian estimation.

Since the complete likelihood is given by (39), the posterior distribution of the  $N_{it}$  is identical to that in the case of stratified sampling (section 4) noting that  $n_i = 0$  is allowed for the latter. Therefore, the 'Bayes estimator' of  $\bar{Y}$  is given by (25) and it may be recast as

$$E'(\bar{Y}) = N^{-1} \sum_1 N_i \left[ \left( 1 - \frac{n_i}{N_i} \right) \sum_t \frac{n_{it} + v_{it}}{n_i + v_i} y_{it} + \frac{n_i}{N_i} \sum_t \frac{v_{it}}{v_i} y_{it} \right] \\ + N^{-1} \sum_2 N_i \sum_t \frac{v_{it}}{v_i} y_{it} \quad (42)$$

where  $\Sigma_1$  and  $\Sigma_2$  respectively the summations over sampled and non-sampled primaries. It should be noted that we must have a prior distribution from each primary. If the prior distribution is solely based on pilot samples, this implies that the pilot sample must include at least one secondary unit from each primary.

The above analysis clearly shows that the sampling procedure adopted for selection of the primaries is entirely irrelevant as far as a full Bayesian analysis is concerned. However, if the likelihood based on a selected estimator is used for a (partial) Bayesian analysis based on insufficient statistics, then the posterior distribution and, hence, the 'Bayes estimator' would depend on the sampling procedure. These are the two alternatives available to the Bayesian analyst and, at this stage, we do not wish to take sides in this issue.

#### REFERENCES

- Draper, N. R. and Guttman, I., "Some Bayesian stratified two-phase sampling results", Biometrika, 55 (1968), in press.
- Ericson, W. A., "Optimum stratified sampling using prior information", J. Amer. Statist. Assoc., 60 (1965), 750-71.
- Godambe, V. P., "A new approach to sampling from finite populations. II.", J. Roy. Statist. Soc., B, 28 (1966), 320-8.
- Hartley, H. O. and Rao, J.N.K., "A new estimation theory for sample surveys", Biometrika, 55 (1968), in press.
- Hoadley, A. B., "Properties of the compound multinomial distribution useful in a Bayesian analysis of categorical data from finite populations", Submitted for publication (manuscript seen by courtesy of the author).
- Kullback, S., "On certain distributions derived from the multinomial distribution", Ann. Math. Statist., 8 (1937), 127-44.
- Rao, J.N.K., "Bayesian optimisation of stratified sampling", submitted for publication.