The Annals of Mathematical Statistics
1971, Vol. 42, No. 4, 1425–1428

SHORT COMMUNICATIONS

A NOTE ON ADMISSIBLE SAMPLING DESIGNS FOR A FINITE POPULATION

BY V. M. JOSHI

Secretary, Maharashtra Government, Bombay

1. Preliminary. Let U be a finite population of units $u_1, u_2, \cdots, u_N$. A sample s means any non-empty subset of U. A sampling design P is determined by defining a probability P on the set S of all possible samples s, $P(s)$ denoting the probability of the sample s. With each unit u_i is associated a variate value x_i, $i = 1, 2, \cdots, N$. $\mathbf{x} = (x_1, x_2, \cdots, x_N)$ denotes a point in the N-space R_N. Then for estimating the population total

$$T(\mathbf{x}) = \textstyle\sum_{i=1}^{N} x_i \tag{1}$$

the Horvitz–Thompson estimate (H–T estimate for short) is given by

$$\bar{e}(s, \mathbf{x}) = \sum_{i \in s} \frac{x_i}{\pi_i} \tag{2}$$

For unbiased estimation of $T(\mathbf{x})$ to be possible, it is a necessary condition that $\pi_i > 0$, $i = 1, 2, \cdots, N$. Throughout the following we restrict ourselves to the class C of sampling designs, for which this condition is satisfied and admissibility of a sampling design P means admissibility within the class C.

The variance of the H–T estimate is given by

$$V(\bar{e}, \mathbf{x}) = \sum_{i=1}^{N} \frac{x_i^2}{\pi_i} + 2 \sum_{1 \leqq i < j \leqq N} \frac{\pi_{ij}}{\pi_i \pi_j} x_i x_j - T^2(\mathbf{x}). \tag{3}$$

In (2) and (3), π_i and π_{ij} are respectively the inclusion probabilities of the units u_i and the pair of units u_i, u_j, i.e.

$$\begin{aligned} \pi_i &= \textstyle\sum_{s \ni i} P(s), \\ \pi_{ij} &= \textstyle\sum_{s \ni i,j} P(s), \qquad i, j = 1, 2, \cdots, N. \end{aligned} \tag{4}$$

In (2), (3) and (4) we have written $i \in s$ for $u_i \in s$, and similarly for $s \ni i$ and $s \ni i, j$.

The expected sample size for a given sampling design P is given by

$$v = \textstyle\sum_{s \in s} P_s n(s), \tag{5}$$

where $n(s)$ denotes the size of the sample s, i.e. the number of units u_i which belong to s.

Let P' be another sampling design and for P' let $V'(\bar{e}, \mathbf{x})$ and s' be the variance of the H–T estimate and the expected sample size. Suppose the sampling cost is

Received November 4, 1969; revised November 17, 1970.

the same per unit and the variance is taken as loss function. Then if the H–T estimate is used for estimating the population total, the sampling design P' is uniformly superior to P if,

$$(6)\qquad \text{(i)}\quad v' \leqq v, \qquad \text{and}$$
$$\text{(ii)}\quad V'(\bar{e}, \mathbf{x}) \leqq V(\bar{e}, \mathbf{x}) \qquad \text{for all } \mathbf{x} \in R_N,$$

and the strict inequality holds either in (6)–(i) or for at least one $\mathbf{x} \in R_N$ in (6)–(ii).

A sampling design P is admissible if there exists no P' uniformly superior to it. With this definition, it was shown by Godambe and the author (1965) that every sampling design of fixed sample size, i.e. for which $P(s) = 0$ unless $n(s) =$ some fixed number m, is admissible.

Now in survey sampling, it is generally the case that the variate assumes only nonnegative values. When this is the case, we may replace in (6)–(ii) R_N by its positive quadrant $R_N{}^+$. It is suggested by Hanurav (1968) that with this restriction a stronger result may hold, and for integral s, sampling designs of fixed size v may be the only admissible designs with a corresponding result for non-integral v.

Suppose that for $\mathbf{x} \in R_N{}^+$, a sampling design P' uniformly superior to P exists, and let w_i' and π_{ij}' be the inclusion probabilities for P' of u_i and the pair u_i, u_j. Considering a point $\mathbf{x}$, for which $x_i \neq 0$ for some index i and $x_i = 0$ for all $j \neq i, j = 1, 2, \cdots, N$, it is seen from (3) that (6)–(ii) implies that

$$(7)\qquad \pi_i' \geqq \pi_i, \qquad i = 1, 2, \cdots, N.$$

By the well-known relation between s and π_i,

$$(8)\qquad \begin{aligned} \textstyle\sum_{i=1}^N \pi_i' &= v' \\ &\leqq v \qquad \text{by (6)–(i)} \\ &= \textstyle\sum_{i=1}^N \pi_i. \end{aligned}$$

Equations (7) and (8) together give,

$$(9)\qquad \pi_i' = \pi_i, \qquad i = 1, 2, \cdots, N,$$

and hence by (6)–(ii), since $x_i x_j \geqq 0$,

$$(10)\qquad \pi_{ij}' \leqq \pi_{ij}, \qquad i, j = 1, 2, \cdots, N, i \neq j,$$

with the strict inequality holding in (10) for at least one pair i, j.

Let $V(n(s))$ and $V'(n(s))$ respectively be the variances of the sample size for the sampling designs P and P'. Again, by a well-known relation,

$$(11)\qquad \textstyle\sum_{1 \leqq i < j \leqq N} \pi_{ij} = \frac{1}{2}v(v-1) + \frac{1}{2}V(n(s)),$$

with a corresponding expression for $V'(n(s))$. Then (10) leads Hanurav (1968) to suggest that perhaps the only admissible sampling designs are those which for given v minimize the variance $V(n(s))$. This means that for integral v the only admissible sampling designs will be those of fixed sample size v for which $V(n(s))$

assumes its minimum value 0, and more generally for non-integral v, the only admissible sampling designs will be those which assign positive probability only to samples of sizes $[v]$ and $[v]+1$, where $[v]$ is the integral part of v, as the variance $V(n(s))$ attains its minimum value for such sampling designs, the minimum value being $\theta(1-\theta)$ where $\theta = v-[v]$. In the following section we show that this conjecture is not true.

Hanurav (1968) also raises the question of determining the minimal complete class of sampling designs. In the following section we obtain a complete (but not the minimal complete) class.

2. A necessary condition for admissibility. A necessary condition for a sampling design being admissible when $\mathbf{x}$ is restricted to $R_N{}^+$ is provided by the following:

PROPOSITION 2.1. *A sampling design P is inadmissible when* $\mathbf{x}$ *is restricted to* $R_N{}^+$, *if there exists at least one pair of samples s_1 and s_2, such that*

$$\text{(i)}\quad P(s_1) > 0, P(s_2) > 0,$$

$$\text{(ii)}\quad s_1 \subset s_2, \tag{12}$$

and

$$\text{(iii)}\quad n(s_2) \geqq n(s_1)+2.$$

PROOF. By (12)–(iii), there exist at least two units, say u_k, u_l, such that $u_k, u_l \in s_2$ and $u_k, u_l \notin s_1$. Let

$$\text{(i) } s_3 = s_1+u_k, \qquad \text{(ii) } s_4 = s_2-u_k \qquad \text{and} \tag{13}$$
$$\text{(iii) } p = \min[P(s_1), P(s_2)] > 0 \qquad \text{by (12)–(i).}$$

We define an alternative sampling design P' by

$$P'(s_1) = P(s_1)-p, \qquad P'(s_2) = P(s_2)-p, \qquad P'(s_3) = P(s_3)+p, \tag{14}$$
$$P'(s_4) = P(s_4)+p, \qquad \text{and } P'(s) = P(s), \qquad \text{if } s \neq s_1, s_2, s_3, \quad \text{or } s_4.$$

It is easily verified using (13)–(iii) that P' is a probability on S, that the probabilities π_i' are the same as π_i, $i = 1, 2, \cdots, N$ and $\pi_{ij}' = \pi_{ij}$ for all (i, j), except that for all l such that $u_l \in (s_2-s_1)$,

$$\pi_{kl}' = \pi_{kl}-p < \pi_{kl} \quad \text{as} \quad p > 0 \quad \text{by (12)–(iii).}$$

Thus (9) and (10) are satisfied for P' and hence P is inadmissible.

An application of Proposition 2.1 is the following:

APPLICATION 2.1. Any sampling design which assigns positive probability to samples of at least two different sizes m_1 and m_2, with $m_2 \geqq m_1+2$, the samples of sizes m_1 and m_2 being drawn by simple random sampling is inadmissible.

By Proposition 2.1 we determine a complete class of sampling designs, namely the class C for which the conditions in (12) are not satisfied. This class is not, however, minimal complete, i.e. that (12) should not hold is not sufficient for admissibility, as is seen from the following example.

EXAMPLE 2.1. The population U consists of eight units, $u_1, u_2, \cdots, u_8$. The sampling design P assigns positive probability of $\frac{1}{4}$ each, to only the four samples

$$s_1 = (1,2,3), \quad s_2 = (1,2,4,5,6), \quad s_3 = (6,7,8) \text{ and } s_4 = (3,4,5,7,8).$$

The sampling design P' assigns probability of $\frac{1}{4}$ each to only the four samples

$$s_1' = (1,2,3,4); \quad s_2' = (1,2,5,6), \quad s_3' = (3,5,7,8) \text{ and}$$

$$s_4' = (4,6,7,8). \quad \text{It is easily verified that}$$

$$\pi_i' = \pi_i = \tfrac{1}{2}, \quad i = 1,2,\cdots,8.$$

$$\pi'_{45} = 0 < \pi_{45} = \tfrac{1}{2}$$

and $$\pi'_{ij} = \pi_{ij} \quad \text{for all other pairs } (i, j).$$

Thus P is not admissible, though (12) does not hold for P.

Lastly we give an example of a class of admissible sampling designs:

PROPOSITION 2.2. *Every sampling design P which is such that every sample s_1 for which $P(s_1) > 0$, includes at least one unit u_i, say, which does not belong to any other s_2 for which $P(s_2) > 0$, is admissible.*

The proof is obvious.

As a special case of Proposition 2.2 it follows that all sampling designs designated as "unicluster designs" by Hanurav (1968) (viz. sampling designs such that for any pair of samples s_1 and s_2, if both $P(s_1) > 0$ and $P(s_2) > 0$ then s_1 and s_2, are disjoint) are admissible.

The falsity of the conjecture mentioned at the end of Section 1 follows from Proposition 2.2. For example, suppose the population U consists of five units $u_1, u_2, \cdots, u_5$. Let $s_1 = (u_1, u_2)$, $s_2 = (u_1, u_3, u_4, u_5)$; and let P be such that $P(s_1) = P(s_2) = \frac{1}{2}$, $P(s) = 0$, for $s \neq s_1$ or s_2. Then P is admissible though the variance is minimized only for a sampling design of fixed size 3.

REFERENCES

[1] GODAMBE, V. P. and JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations—I. *Ann. Math. Statist.* **36** 1707–1722.

[2] HANURAV, T. V. (1968). Hyperadmissibility and optimum estimators for sampling finite populations. *Ann. Math. Statist.* **39** 621–641.