The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 278–281

NOTE ON A MINIMAX DESIGN FOR CLUSTER SAMPLING

BY V. M. JOSHI

Maharashtra Government, Bombay

1. Introduction. The problem of determining the minimax procedure for estimating the population mean, with two-stage cluster sampling has been recently discussed by Aggarwal (1966), without however giving a general solution. A general solution of the problem is presented in this note.

2. Preliminary. We shall refer to Aggarwal's paper (1966) as the 'Main Paper' or shortly as M and use throughout the same notation as in M. It is shown in equation (6.1) in Section 6 of M, that with given m clusters, the minimax sampling scheme is obtained by choosing the n_i, $i = 1, 2, \cdots, m$, so as to minimize the risk,

$$R(\mu, \delta^*) = \{\textstyle\sum_{i=1}^m n_i/(n_i\sigma_b^2 + \sigma_i^2)\}^{-1} + c_b m + \textstyle\sum_{i=1}^m n_i c_i. \quad (1)$$

It is further observed in M, that theoretically speaking the risk (1) should be minimized over the choice of n_i, under the restriction that they be positive integers, but that even without this restriction it does not seem possible to solve the problem of minimizing (1), in general; and that it may be possible only to obtain approximate solutions under some simplifying assumptions. The solution for one such particular case is derived in Section 9 of M.

In the following we obtain a general solution giving non-negative values of n_i, which minimize right hand side of (1), provided the restriction to integral values is ignored.

3. Main result. We put,

$$S = \textstyle\sum_{i=1}^m n_i/(n_i\sigma_b^2 + \sigma_i^2), \quad (2)$$

$$R = R(\mu, \delta^*). \quad (3)$$

We are concerned with only the positive quadrant of the m-spcce of the variables n_i, $i = 1, 2, \cdots, m$, defined by $n_i \geqq 0$. We shall refer to this space as the n_i-space. Suppose R has a minimum (in the calculus sense) in the positive quadrant of the n_i-space. Then at the minimum point we must have

$$\partial R/\partial n_i = 0, \qquad i = 1, 2, \cdots, m.$$

By differentiation, we obtain from (1)

$$\partial R/\partial n_i = -\sigma i^2/S^2(n_i\sigma_b^2 + \sigma_i^2)^2 + c_i, \quad (4)$$

so that by equating $\partial R/\partial n_i$ to 0, we have

$$n_i\sigma_b^2 + \sigma i^2 = \sigma i/Sc_i^{\frac{1}{2}}, \quad (5)$$

Received 22 December 1966.

and,

$$n_i\sigma_b^2 = \sigma i/Sc_i^{\frac{1}{2}} - \sigma_i^2. \tag{6}$$

Dividing (6) by (5), and summing up from $i = 1$ to m, we get, using (2),

$$S \cdot \sigma_b^2 = m - S \sum_{i=1}^m c_i^{\frac{1}{2}}\sigma_i ,$$

i.e.

$$S^{-1} = m^{-1}\sigma_b^2 + m^{-1} \sum_{i=1}^m c_i^{\frac{1}{2}}\sigma_i . \tag{7}$$

Substituting (7) in (6), we get

$$n_i\sigma_b^2 = \sigma_i c_i^{-\frac{1}{2}}\{m^{-1}\sigma_b^2 + m^{-1} \sum_{i=1}^m c_i\sigma_i - \sigma_i c_i^{\frac{1}{2}}\}. \tag{8}$$

Since by assumption, this minimum occurs in the positive quadrant of the n_i-space, we must have

$$m^{-1}\sigma_b^2 + m^{-1} \sum_{i=1}^m c_i^{\frac{1}{2}}\sigma_i \geqq \sigma_i c_i^{\frac{1}{2}} \qquad \text{for } i = 1, 2, \cdots, m. \tag{9}$$

Thus a minimum of R occurs in the positive quadrant, only if (9) is satisfied, and it is seen that if (9) is satisfied the values of n_i, given by (8) are all non-negative and as they are uniquely determined they must yield the minimum value of R, which by assumption exists in the positive quadrant.

Next suppose that (9) is not satisfied Then R has no minimum in the calculus sense at any point of the positive quadrant. It may however have a minimum value on a point on the boundary of the positive quadrant, the minimum being in the sense of the least value in the part of the neighbourhood of the boundary point which falls within the positive quadrant of the n_i-space. Such a minimum occurs on a boundary point, if conditions such as the following are satisfied:

$$\begin{aligned} n_i &= 0, && i = i_1, i_2, \cdots, i_k, \\ n_i &\geqq 0, && \text{for other } i, \\ \partial R/\partial n_i &> 0, && i = i_1, i_2, \cdots, i_k, \\ \partial R/\partial n_i &= 0, && \text{for other } i. \end{aligned} \tag{10}$$

We shall investigate the conditions under which (10) can hold. From (4), it is seen that $\partial R/\partial n_i > 0$, if and only if,

$$S^{-1} < c_i^{\frac{1}{2}}\sigma_i . \tag{11}$$

Without loss of generality, we can assume that the clusters are numbered in such a way that the numbers $\{c_i^{\frac{1}{2}}\sigma_i\}$ form a non-decreasing series. It then follows from (11), the indices i for which $\partial R/\partial n_i > 0$ in (10), must be the highest indices, i.e. they are given by $i = k + 1, k + 2, \cdots, m$, for some k.

We therefore assume that in (10),

$$n_i = 0 \quad \text{for} \quad i = k + 1, \cdots, m. \tag{12}$$

Then dividing (6) by (5), and summing from $i = 1, 2, \cdots, k$, and noting that the remaining terms in the summation for S in the right hand side of (2) vanish, we have

$$S \cdot \sigma_b^2 = k - S \cdot \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i ,$$

so that,

$$S^{-1} = k^{-1} \sigma_b^2 + k^{-1} \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i . \tag{13}$$

Then by substituting (13) in (6), we get

$$n_i \sigma_b^2 = \sigma_i c_i^{-\frac{1}{2}} \{ k^{-1} \sigma_b^2 + k^{-1} \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i - c_i^{\frac{1}{2}} \sigma_i \}, \qquad i = 1, 2, \cdots, k. \tag{14}$$

From (13) and (14) it follows that, in order to satisfy (10), we must have

$$\begin{aligned} \sigma_b^2 + \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i &\geqq k c_i^{\frac{1}{2}} \sigma_i , \qquad i = 1, 2, \cdots, k; \\ \sigma_b^2 + \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i &< k c_i^{\frac{1}{2}} \sigma_i , \qquad i = k+1, \cdots, m. \end{aligned} \tag{15}$$

We shall next show that there always exists one, and only one value of k, for which (15) is satisfied.

Since the numbers $\{c_i^{\frac{1}{2}} \sigma_i\}$ are non-decreasing, the inequalities (15), are equivalent to

$$\sigma_b^2 + \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i - k c_k^{\frac{1}{2}} \sigma_k \geqq 0, \tag{16}$$

$$\sigma_b^2 + \sum_{i=1}^{k} c_i^{\frac{1}{2}} \sigma_i - k c_{k+1}^{\frac{1}{2}} \sigma_{k+1} < 0. \tag{17}$$

Putting $k = 1$, it is seen that (16) is satisfied. Further, since by assumption (9) does not hold, (16) does not hold for $k = m$. Hence assigning to k successively the values 1, 2, we must reach some lowest value k for which (16) and (17) hold. We now show that there is no other such value. Let $k' > k$. Applying (17) and the monotonicity of $c_i^{\frac{1}{2}} \cdot \sigma_i$ we have

$$\begin{aligned} \sigma_b^2 + \sum_{i=1}^{k'} c_i^{\frac{1}{2}} \cdot \sigma_i = \sigma_b^2 + \sum_{i=1}^{k} c_i^{\frac{1}{2}} \cdot \sigma_i + \sum_{k+1}^{k'} c_i^{\frac{1}{2}} \cdot \sigma_i &< k c_{k+1}^{\frac{1}{2}} \cdot \sigma_{k+1} \\ &+ (k' - k) c_{k'}^{\frac{1}{2}} \cdot \sigma_{k'} \leqq k' c_{k'}^{\frac{1}{2}} \cdot \sigma_{k'}. \end{aligned}$$

Hence (16) cannot apply for k'.

Using this value of k, we obtain by (12) and (14) the unique set of values of n_i, for which the relations in (10) are satisfied.

By elementary considerations it is seen that R must have a least value in the positive quadrant of the n_i-space; since by assumption, R has no minimum in the calculus sense, at this least value, relations like those in (10) must hold. But as shown above, there is only a unique set of values of n_i, satisfying (10). This unique set accordingly gives the least value of R in the positive quadrant of the n_i-space.

The values of n_i given by (8) are the same as those given by (14), where k is taken equal to m. Hence the rule for determining the minimizing values of n_i

may be stated as follows: construct successively the series of numbers,

$$t_k = k^{-1}\sigma_b^2 + k^{-1} \textstyle\sum_{i=1}^{k} c_i^{\frac{1}{2}} \cdot \sigma_i \tag{18}$$

for $k = 1, 2, \cdots$; continue the series until we reach a value of k, for which (16) and (17) hold, or until $k = m$. This procedure always gives a unique value of k. Using this value in (13) and (12), we get the minimizing values of n_i.

4. An application. As an application of our formulae, we shall verify that the values of n_i given by our formulae agree with those obtained for the particular case in Section 9 of M. In this case, it is assumed that

$$\sigma_i \cdot c_i^{\frac{1}{2}} = c^{\frac{1}{2}} \qquad \text{for all} \quad i.$$

Clearly condition (9) is satisfied. Hence by (8),

$$n_i = m^{-1}\sigma_i / c_i^{\frac{1}{2}} = m^{-1} c^{\frac{1}{2}} / c_i, \qquad i = 1, 2, \cdots, m,$$

which are the values derived in Section 9 of M.

5. Integral values. Actually the values of n_i must be integral. The integral values may be obtained by rounding to the nearest integer, the values given by (8) or (14). These nearest integral values, would not necessarily give the minimax design and some adjustments of the values by trial and error may be necessary. In most cases however the nearest integral values will give at least a design close to the minimax design.

Acknowledgment. I am very grateful to the referee for his valuable suggestions.

REFERENCE

[1] AGGARWAL, OM. P. (1966). Bayes and minimax procedures for estimating the arithmetic mean of a population with two-stage sampling. *Ann. Math. Statist.* **37** 1186–1195.