ADMISSIBILITY AND BAYES ESTIMATION IN SAMPLING FINITE POPULATIONS—IV

By V. M. Joshi

Institute of Science, Bombay, and University of North Carolina

1. Introduction. This paper, which we call Part IV, is a continuation of the previous papers on the same subject, Part I by Joshi and Godambe (1965) and Parts II and III by Joshi (1965). In the introductory section of Part III a note was added at the proof stage, that it was realized that the result proved in that part relating to the admissibility of the Horvitz-Thomson estimate was valid for the class of all measurable estimates, without any 'regularity' restrictions on the class. We give a clarification of this point here and also add a supplementary, though minor result, subsequently obtained, that even the measurability restriction is removed for the special cases of sample size $m = 1$ and 2.

Next we give new results obtained mainly by applying the method developed in Part III, regarding the admissibility of the well-known ratio estimates, and the regression estimate for finite populations. While one ratio estimate is found to be always admissible whatever the sampling design, in the class of all measurable estimates, the other ratio estimate is shown to be necessarily admissible only when the sampling design is of fixed size, and that too subject to a certain condition. The regression estimate is also shown to be not always admissible.

2. Notation. The same notation is followed as in the previous parts, as specified in Section 2, Part I and Section 2, Part II. The definitions and preliminaries in Section 2 of Part I also all apply.

3. Superfluity of regularity conditions. The question relates to (15) in Part III where the Cramér-Rao lower bound for the variance of an estimate is assumed to apply. By (14) in Part III, the probability density is that of m independently distributed normal variates. The Wolfowitz conditions (1947) which are sufficient for the Cramér-Rao inequality are considered and shown to be all satisfied for a normal density function in Problem 1 of Hodges and Lehmann (1951). Our case is slightly different in that the variances of the variates may be unequal. But it is easily seen that this makes no difference in regard to conditions (i) to (iv) stated in Problem 1 of Hodges and Lehmann (1951). It will therefore suffice to verify that the remaining condition (v) is also satisfied. For the density function $L(y/\theta)$ in (12) of Part III, this condition becomes:

(v) The expression $\int g(s, y)L(y/\theta)\,dy$ may be differentiated under the integral sign.

But this follows readily from the fact that by definition of $g(s, y)$ in (5) and (10) Part III, $[g(s, y)]^{\frac{1}{2}}$ is bounded for all y, by a quadratic expression in y, say $u(y)$, i.e.

Received 30 November 1965; revised 22 June 1966.

$$(g(s, y))^2 \leqq u(y). \tag{3.1}$$

Now writing L for short for $L(y/\theta)$, and denoting corresponding increments of θ and $L^{-1}(\partial L/\partial\theta)$ by $\Delta\theta$ and $\Delta(L^{-1}(\partial L/\partial\theta))$, it is seen from (12) in Part III that

$$\begin{aligned} |\Delta(L^{-1}(\partial L/\partial\theta))/\Delta\theta| &= |\textstyle\sum_{r=1}^{N}[(y_r - \theta) - \Delta\theta]/[1 - (br)^{-1}]| \\ &\leqq \textstyle\sum_{r=1}^{N}|y_r - \theta|/[1 - (br)^{-1}] \\ &\quad + |\Delta\theta|\cdot\textstyle\sum[1 - (br)^{-1}]^{-1} \leqq v(y, \theta, \Delta\theta) \quad \text{say.} \end{aligned} \tag{3.2}$$

Then by (3.1) and (3.2)

$$\int |g(s, y)\Delta(L^{-1}(\partial L/\partial\theta))/\Delta\theta| L\, dy \leqq \int [u(y)]^{\frac{1}{2}} v(y, \theta, \Delta\theta) L\, dy < \infty. \tag{3.3}$$

(v) then follows from (3.3) by the dominated convergence theorem, see for example Loève (1960). Thus no other regularity conditions are required.

4. Special cases $m = 1$ and $m = 2$. In these cases the strict admissibility of the estimate $\bar{e}(s, x)$ in Theorem 3.1 of Part III, can be proved directly by an algebraic method, so that no measurability restriction applies. Putting $e'(s, x) = \bar{e}(s, x) + h(s, x)$ (2) in Part III, reduces to

$$\textstyle\sum_{s\varepsilon\bar{S}} p(s)h^2(s, x) + 2\sum_{s\varepsilon\bar{S}} p(s)h(s, x)[\bar{e}(s, x) - T(x)] \leqq 0. \tag{4.1}$$

Here $\bar{S}$ is the set of all samples s for which $p(s) \neq 0$.

Case $m = 1$. Let $h_i(t)$ denote the value of $h(s, x)$ for the sample s_i which consists of unit u_i, $i = 1, 2, \cdots, N$. Consider (4.1) at the point $P_t \,\varepsilon\, R_N$ with co-ordinates $x_r = t/b_r$, $r = 1, 2, \cdots, N$. Then by condition (ii) of Theorem 3.1 in Part III, at the point P_t, $T(x) = t = \bar{e}(s_i, x)$ for all $s_i \,\varepsilon\, \bar{S}$ and hence from (4.1), $\sum_{s_i\varepsilon\bar{S}} p(s_i)h_i^2(t) = 0$ so that $h_i(t) = 0$ for all $s_i = \bar{S}$, and since t is arbitrary it follows that $h_i(t) = h(s_i, x) = 0$ for all $x \,\varepsilon\, R_n$ and all $s \,\varepsilon\, \bar{S}$. The admissibility of $\bar{e}(s, x)$ follows from this.

Case $m = 2$. Denote by s_{ij}, with $i < j$, the sample consisting of units u_i, u_j and let $h_{ij}(t_1, t_2)$ denote the value of $h(s_{ij}, x)$ when $x_i = t_1/b_j$ and $x_j = t_2/b_j$, $ij = 1, 2, \cdots, N$. Then again by considering (4.1) at the point P_i with co-ordinates $x_r = t/b_r$, $r = 1, 2, \cdots, N$, it is seen as in the Case $m = 1$, that $h_{ij}(t, t) = 0$ for all $s_{ij} \,\varepsilon\, \bar{S}$. Next let $t_1 \neq t_2$ and consider the set of $2N$ points P_r, Q_r, $(r = 1, 2, \cdots, N)$ such that for P_r, only the co-ordinate $x_r = t_1/b_r$ and all other co-ordinates $x_s = t_2/b_s$, $S \neq r$, and similarly for Q_r only $x_r = t_2/b_r$ and all other co-ordinates $= t_1/b_s$. Then at P_r and Q_r, $h(s, x) = 0$ for every s which does not contain the unit u_r. For a sample s containing units u_r and u_s say, at the point P_r

$$\begin{aligned} \bar{e}(s, x) - T(x) &= t_1 + t_2 - t_1/b_r - t_2\cdot\textstyle\sum_{j=1, j\neq r}^{N} b_j^{-1} \\ &= (t_1 - t_2)\cdot(1 - b_r^{-1}). \end{aligned} \tag{4.2}$$

Since by condition (ii) of Theorem 3.1, Part III, $\sum_{r=1}^{N} b_r^{-1} = 2$ for $m = 2$. Similarly at Q_r,

$$\bar{e}(s, x) - T(x) = (t_2 - t_1)(1 - b_r^{-1}). \tag{4.3}$$

Note that by condition (i) of Theorem 3.1 of Part III, the factor $1 - b_r^{-1} > 0$ for all $r = 1, 2, \cdots, N$. Now multiply the relation (4.1) at points P_r and Q_r by $(1 - b_r)^{-1}$ and sum over all the $2N$ points $P_1, \cdots, P_N, Q_1, \cdots, Q_N$. In the sum of the 2nd term in the left hand side of (4.1), the term $h_{ij}(t_1, t_2)$ occurs with coefficient $(t_1 - t_2)$ at P_1 and with coefficient $(t_2 - t_1)$ at Q_j and hence all such terms cancel out, giving $\sum_{s_{ij} \varepsilon \bar{S}} p(s_{ij}) h_{ij}^2(t_1, t_2) \leqq 0$ which implies $h_{ij}(t_1, t_2) = 0$ for all $s_{ij} \varepsilon \bar{S}$. Since t_1, t_2 are arbitrary, and since $h_{ij}(t, t)$ also vanishes, it follows that $h(s_{ij}, x) = 0$ for all $s_{ij} \varepsilon \bar{S}$, and all $x \varepsilon R_N$, from which the admissibility of $\bar{e}(s, x)$ for $m = 2$ follows.

For $m \geqq 3$, the measurability restriction remains, which however seems immaterial as no non-measurable functions can be used as estimates in practice. An example of a non-measurable set due to Vitali is given by Rogosinski (1952), p. 73. The set V is formed by taking from each set V_ξ of all numbers $\xi + r$ where ξ is an irrational number, and r runs through the set of all rationals, some number say the 1st according to a given enumeration of the rationals which falls in $[0, \frac{1}{2}]$. A non-measurable function based on the set may be defined by putting $\phi(x) = 1$ if $x \varepsilon V$ and $\phi(x) = 0$ for $x \not\varepsilon V$. Through the function ϕ is thus formally defined, we cannot state its value for any given point say $x = 3^{-\frac{1}{2}}$ as whether $3^{-\frac{1}{2}} \varepsilon V$ or not, depends on which of the numbers $3^{-\frac{1}{2}} + \xi$ was taken as the 1sf number in forming V_ξ. It is stated by Rogosinski that all known examples ot non-measurable sets are based on the axiom of choice. It therefore seems that though non-measurable sets and non-measurable functions may for theoretical purposes be defined by using the axiom of choice, the non-measurable functions cannot be used as estimates in practice.

5. Admissibility of a ratio estimate. $y_1 > 0$, $i = 1, 2, \cdots, N$, are known positive constants and $Y = \sum_{i=1}^{N} y_i$. We shall show that the ratio estimate

$$e_R(s, x) = (\textstyle\sum_{r \varepsilon s} x_r / \sum_{r \varepsilon s} y_r) \cdot Y \tag{5.1}$$

is admissible in the class of all measurable estimates. Suppose it is not, then there exists an estimate $e'(s, x)$ satisfying

$$\textstyle\sum_{s \varepsilon \bar{S}} p(s)[e'(s, x) - T(x)]^2 \leqq \sum_{s \varepsilon \bar{S}} p(s)[e_R(s, x) - T(x)]^2 \tag{5.2}$$

with the strict inequality holding for some $x \varepsilon R_N$. Take the expectations of both sides of (5.2), with respect to a prior distribution, such that the variates x_i are all distributed independently with $E(x_i) = \theta \cdot y_i$. Put

$$\begin{aligned} A(s) &= \textstyle\sum_{r \not\varepsilon s} y_r ; \\ g(s, x) &= [A(s)]^{-1}[e'(s, x) - \textstyle\sum_{r \varepsilon s} x_r]; \\ \bar{x}(s) &= (\textstyle\sum_{r \varepsilon s} x_r / \sum_{r \varepsilon s} y_r). \end{aligned} \tag{5.3}$$

In the left hand side of (5.2),

$$E[e'(s, x) - T(x)]^2 = E[e'(s, x) - \sum_{r\varepsilon s} x_r - \theta\cdot A(s) - \sum_{r\notin s} (x_r - \theta\cdot y_r)]^2$$
$$= A^2(s)E[g(s, x) - \theta]^2 + \sum_{r\notin s} E(x_r - \theta\cdot y_r)^2,$$

the cross terms vanishing due to independence of the variates. In the right hand side of (5.2), substituting for $e_R(s, x)$ by (5.1),

$$E[(\sum_{r\varepsilon s} x_r/\sum_{r\varepsilon s} y_r)Y - T(x)]^2$$
$$= E[\sum_{r\varepsilon s} x_r(Y/\sum_{r\varepsilon s} y_r - 1) - \theta\cdot A(s) - \sum_{r\notin s} (x_r - \theta\cdot y_r)]^2$$
$$= A^2(s)E(\bar{x}(s) - \theta)^2 + \sum_{r\notin s} E(x_r - \theta\cdot y_r)^2.$$

Hence on taking expectations of both sides of (5.2) and cancelling out common terms, we get

$$(5.4)\quad \sum_{s\varepsilon\bar{S}} p(s)A^2(s)E[g(s, x) - \theta]^2 \leqq \sum_{s\varepsilon\bar{S}} p(s)A^2(s)E[\bar{x}(s) - \theta]^2.$$

We now make the further assumption that each variate x_r, $r = 1, 2, \cdots, N$, is distributed normally with variance proportional to y_r so that

$$(5.5)\quad \sigma_r^2 = k\cdot y_r$$

where k is a constant >0. Then

$$(5.6)\quad E[\bar{x}(s) - \theta]^2 = k/\sum_{r\varepsilon s} y_r = k/B(s)$$

where

$$(5.7)\quad B(s) = \sum_{r\varepsilon s} y_r.$$

The frequency function for the variates x_r for which $r \varepsilon s$, is

$$L = (2\pi)^{-n(s)/2} \prod_{r\varepsilon s} \sigma_r^{-1} \exp[-\tfrac{1}{2} \sum_{r\varepsilon s} (x_r - \theta y_r)^2/\sigma_r^2].$$

Hence

$$(5.8)\quad E(\partial \log L/\partial\theta)^2 = E[\sum_{r\varepsilon s} (y_r/\sigma_r^2)(x_r - \theta\cdot y_r)]^2$$
$$= B(s)/k \qquad (\text{by } (5.5) \text{ and } (5.7)).$$

Next putting $E(g(s, x)) = \theta + b(s, \theta)$ and using the Cramér-Rao inequality as in (15) of Part III, we get from (5.4)

$$\sum_{s\varepsilon\bar{S}} p(s)A^2(s)b^2(s, \theta) + k\sum p(s)[A^2(s)/B(s)](1 + b'(s, \theta))^2$$
$$\leqq k\sum_{s\varepsilon\bar{S}} p(s)A^2(s)/B(s).$$

We now define the weighted mean bias $\bar{b}(\theta)$ by

$$\bar{b}(\theta) = \sum_{s\varepsilon\bar{S}} p(s)[A^2(s)/B(s)]\cdot b(s, \theta)/\sum_{s\varepsilon\bar{S}} p(s)A^2(s)/B(s)$$

and proceeding as from (17) to (19) of Part III we get in place of (19) in Part III, (5.9) $$[\sum_{s\varepsilon\bar{S}} p(s)[A^2(s)/B(s)]]^{-1}\cdot k^{-1}\cdot \sum_{s\varepsilon\bar{S}} p(s)\cdot A^2(s)\cdot b^2(s,\theta) + (1+\bar{b}'(s,\theta))^2 \leqq 1.$$

By a subsequent argument, exactly the same as that following (19) in Part III, we get $b(s, \theta) = 0$ for all $s \varepsilon \bar{S}$, so that $g(s, x)$ and $\bar{x}(s)$ being efficient unbiassed estimates of θ, are equal a.e. in R_N, from which follows using (5.3) and (5.1) that $e'(s, x) = e_R(s, x)$ a.e. in R_N. Thus Theorem 3.1 of Part III, (with the necessary obvious verbal modifications) holds good for the estimate $e_R(s, x)$.

We next show that Theorem 4.1 of Part III also holds good for $e_R(s, x)$. Let the hyperplanes Q_{N-k}^{α}, $Q_{N-k}^{\alpha'}$ and the set of samples $\bar{S}_k$ be defined as in Theorem 4.1 of Part III. We establish a 1-1 correspondence between the points of Q_{N-k}^{α} and $Q_{N-k}^{\alpha'}$ by putting $x_r' = x_r + h\cdot y_r$. We shall show that the constant h can be so fixed that for all $s \varepsilon \bar{S}_k$,

(5.10) $$e_R(s, x') - T(x') = e_R(s, x) - T(x).$$

Since every $s \varepsilon \bar{S}_k$ contains each of the last k units,

$$\begin{aligned} e_R(s, x') - T(x') &= [\textstyle\sum_{r=N-k+1}^{N} \alpha_r' + \sum_{r\varepsilon s, r\leqq N-k} x_r + h\sum_{r\varepsilon s, r\leqq N-k} y_r] \\ &\quad \cdot(\textstyle\sum_{r\varepsilon s} y_r)^{-1}\cdot Y - \sum_{r=N-k+1}^{N}\alpha_r' - \sum_{r=1}^{N-k} x_r \\ &\quad - h\cdot \textstyle\sum_{r=1}^{N-k} y_r \\ &= \textstyle\sum_{r=N-k+1}^{N} \alpha_r'(Y/\sum_{r\varepsilon s} y_r - 1) \\ (5.11)\qquad &\quad + [\textstyle\sum_{r\varepsilon s, r\leqq N-k} x_r/\sum_{r\varepsilon s} y_r)\cdot Y - \sum_{r=1}^{N-k} x_r] \\ &\quad + [h(\textstyle\sum_{r\varepsilon s} y_r - \sum_{r=N-k+1}^{N} y_r)/\sum_{r\varepsilon s} y_r]\cdot Y \\ &\quad - h(Y - \textstyle\sum_{r=N-k+1}^{N} y_r) \\ &= \textstyle\sum_{r=N-k+1}^{N} \alpha_r'(Y/\sum_{r\varepsilon s} y_r - 1) - h \\ &\quad \cdot\textstyle\sum_{r=N-k+1}^{N} y_r(Y/\sum_{i\varepsilon s} y_i - 1) \\ &\quad + [(\textstyle\sum_{r\varepsilon s, r\leqq N-k} x_r/\sum_{r\varepsilon s} y_r)\cdot Y - \sum_{r=1}^{N-k} x_r] \end{aligned}$$

and

(5.12) $$T(x) + e_R(s, x) = \textstyle\sum_{r=N-k+1}^{N} \alpha_r(Y/\sum_{i\varepsilon s} y_i - 1) + [(\sum_{i\varepsilon s, i\leqq N-k} x_i/\sum_{i\varepsilon s} y_i)\cdot Y - \sum_{r=1}^{N-k} x_r].$$

Comparing (5.11) and (5.12), (5.10) is seen to be satisfied if

$$h = \textstyle\sum_{r=N-k+1}^{N} \alpha_r' - \sum_{r=N-k+1}^{N} \alpha_r/\sum_{r=N-k+1}^{N} y_r.$$

Hence as in (25) of Part III, the estimate $e'(s, x)$ is extended to every hyperplane $Q_{N-k}^{\alpha'}$, by putting

$$e'(s, x') - T(x') = e'(s, x) - T(x).$$

The whole of the remaining argument in Theorem 4.1 in Part III, now applies without modification and thus Theorem 4.1 in Part III also holds good (with necessary verbal modifications) for $e_R(s, x)$.

Hence Theorem 5.1 in Part III also applies and $h(s, x) = e'(s, x) - e_R(s, x)$ cannot be $\neq 0$ for any point $x \varepsilon R_N$. The strict inequality in (5.2) thus cannot hold at any point. This completes the proof.

It may be noted that in Part III, all the samples $s \varepsilon \bar{S}$, were of fixed size m, while here the size varies. It is however seen that the argument in Theorem 5.1 of Part III is not dependent on the sample size being fixed.

The estimate $e^*(s, x)$ defined in Part II of this paper, is obtained as a particular case of $e_R(s, x)$ by taking all y_r equal.

The result of Part II, is however valid without the restriction of measurability and also holds good for certain subsets of R_N, and hence is not contained as a particular case of the present result.

6. Another ratio estimate. Another common ratio estimate is

$$e_r(s, x) = [Y/n(s)] \textstyle\sum_{r \varepsilon s} x_r/y_r . \tag{6.1}$$

If the sampling design is of fixed size m, i.e. $n(s) = m$ for all $s \varepsilon \bar{S}$, then $e_r(s, x)$ becomes the same as the estimate $\bar{\bar{e}}(s, x)$ defined in $(1)_1$ of Part III, with $b_r = Y/my_r$. Hence the result of that part applies. Condition (ii) of Theorem 3.1 of Part III, viz. $\sum_{r=1}^{N} b_r^{-1} = m$ is always satisfied. Condition (i) is $b_r > 1$, $r = 1, 2, \cdots, N$, which requires that $[\max_r \cdot y_r] < Y/m$. Hence provided this condition is satisfied, the estimate $e_r(s, x)$ is always admissible for fixed sample size design. As shown by example (ii) in Section 5 of Part III, when the condition is not satisfied the estimate may fail to be admissible.

When the sampling design is not of fixed size, the estimate $e_r(s, x)$ may not be admissible as seen from the following example. Population $U = \{u_i, i = 1, 2, \cdots, 9\}$. The sampling design consists of samples of size 3 and 4 all having the same probability; $y_1 = 2$, and $y_i = 1$, for $i = 2, 3, \cdots, 9$; S_{12} denotes the set of the 28 samples each of which contains units u_1 and u_2, and $P^{\alpha\beta}$, the hyperplane in which x_1 and x_2 have fixed values, α and β respectively. Then for $x \varepsilon P^{\alpha\beta}$, in the expression for $\sum_{s \varepsilon S_{12}} [e_r(s, x) - T(x)]^2$, the coefficient of each term $2x_i$, $(i = 3, 4, \cdots, 9)$ is found after reduction, to be $-(11/3)(2\alpha/3 + 7\beta/3) - 6(\alpha/4 + 3\beta/2) = u$ say while the constant term, i.e. independent of x_i, comes to $7(2\alpha/3 + 7\beta/3)^2 + 21(\alpha/4 + 3\beta/2)^2 = V$ say. We now determine constants l, m satisfying

$$-(11/3)l - 6m = u \tag{6.2}$$

and

$$7l^2 + 21m^2 \leqq v. \tag{6.3}$$

This can for example be done by minimizing $7l^2 + 21m^2$, subject to (6.2). The minimizing condition is found to be $l = (11/6)m$ and the constants $(2\alpha/3 + 7\beta/3)$ and $(\alpha/4 + 3\beta/2)$ do not themseleves satisfy this condition except when

$\alpha = 2\beta$, i.e. except when the hyperplane $P^{\alpha\beta}$ contains the st. line $x_i = ky_i$, $i = 1, 2, \cdots, 9$, on which the variance of $e_r(s, x)$ vanishes. Next by putting for $x \varepsilon P^{\alpha\beta}$

$$e'(s, x) = e_r(s, x) + l - (2\alpha/3 + 7\beta/3) \quad \text{if} \quad s \varepsilon S_{12}, \quad \text{and} \quad n(s) = 3,$$

$$e'(s, x) = e_r(s, x) + m - (\alpha/4 + 3\beta/2) \quad \text{if} \quad s \varepsilon S_{12}, \quad \text{and} \quad n(s) = 4,$$

$$e'(s, x) = e_r(s, x) \quad \text{if} \quad s \not\varepsilon S_{12},$$

in $\sum_{s\varepsilon S_{12}} [e'(s, x) - T(x)]^2$, the coefficients of all variable terms involving x_i $(i = 3, \cdots, 9)$ remain unchanged and the constant term is reduced whenever $\alpha \neq 2\beta$, and remains unchanged for $\alpha = 2\beta$. The estimate $e'(s, x)$ thus defined for every hyperplane $P^{\alpha\beta}$, is uniformly superior to $e_r(s, x)$ and has lower variance whenever $\alpha \neq 2\beta$, i.e. a.e. in R_N. Thus in the example, the ratio estimate $e_r(s, x)$ is not even weakly admissible.

The regression estimate is

$$(6.4) \quad e_{\text{reg}}(s, x) = \bar{x}_s + [\textstyle\sum_{i\varepsilon s} (x_i - \bar{x}_s)(y_i - \bar{x}_s)/\sum_{i\varepsilon s} (y_i - \bar{y}_s)^2](Y - \bar{y}_s)$$

where $\bar{x}_s = [n(s)]^{-1} \sum_{i\varepsilon s} x_i$, and $\bar{y}_s = [n(s)]^{-1} \sum_{i\varepsilon s} y_i$.

It is easily verified that an example for the inadmissibility of the estimate $e_{\text{reg}}(s, x)$ can be constructed on similar lines by taking a population $U = (u_i)$, $i = 1, 2, \cdots, 10$ say, $y_1 = 3$, $y_2 = 2$ and $y_i = 1$, for $i = 3, \cdots, 10$; a sampling design consisting of samples of size 4 and 5, all with the same probability and considering the set S_{123} consisting of the 28 samples each of which contains the units u_1, u_2 and u_3 and the hyperplane $P^{\alpha\beta\gamma}$ in which the variates x_1, x_2, x_3 remain fixed, and $= \alpha$, β and γ respectively. It will be found that the variance of the uniformly superior estimate is actually less whenever $\alpha - \beta \neq \beta - \gamma$ i.e. whenever $(\alpha - \beta)/(\beta - \gamma) \neq (y_1 - y_2)/(y_2 - y_3)$, so that the hyperplane $P^{\alpha\beta\gamma}$ does not contain the 2-space generated by the 2 straight lines $x_i = ky_i$, $i = 1, 2, \cdots, 10$, and $x_i = y_i$, $i = 1, \cdots, 10$, in which 2-space the variance of $e_{\text{reg}}(s, x)$ everywhere vanishes. Thus $e'(s, x)$ has lower variance a.e. in R_N.

In the above examples the values of the constants y_i have been suitably adjusted to simplify the arithmetic. But a little consideration shows that it may be possible to construct in more general cases the uniformly superior estimates $e'(s, x)$ by considering the planes $P^{\alpha\beta}$. We take $e'(s, x) = \sum_{r\varepsilon s} \beta_{sr} x_r + \lambda_r$ and determine the coefficients β_{sr}, λ_r so that in the expression $\sum_{s\varepsilon \bar{S}_{12}} p(s)[e'(s, x) - T(x)]^2$ the coefficients of the variable terms x_i and x_j, $(i, j \geqq 3)$ are the same as those in the expression $\sum_{s\varepsilon \bar{S}_{12}} p(s)[e_r(s, x) - T(x)]^2$, while the coefficients for terms x_i^2 and the constant term in the former are $\leqq$ to the corresponding coefficients in the latter, the strict inequality holding in at least one case. For the sampling design considered in the example, the number of constants β_{sr}, λ_r will be seen to be 133, and they will have to satisfy 28 equalities and 7 inequalities so that a solution may exist for more general values of y_i. These remarks apply also to the regression coefficient.

The estimates $e_r(s, x)$ and $e_{\text{reg}}(s, x)$ are however always admissible, whatever the sampling design in the restricted class of all linear estimates. This limited admissibility has however little significance and the proof is also very simple and hence is omitted.

7. Uniform admissibility of the estimate $e^*(s, x)$**.** In Part II of this paper, the estimate $e^*(s, x)$ was proved to be admissible in the entire class of all estimates whatever be the sampling design. We shall now show that this estimate possesses a much stronger 'admissibility' property. As the terms 'weak admissibility' and 'strict admissibility' have already been used with another significance in Part III of this paper, we shall denote this stronger property as that of 'uniform admissibility'. The motivation for extending the concept of admissibility of estimates is as follows:

In this paper so far, admissibility of an estimate has been defined for a particular sampling design d which is taken as given. Let $V(e, d)$ denote the mean squared error of an estimate $e(s, x)$ for a given sampling design d, so that

$$V(e, d) = \sum_{s \varepsilon \bar{S}} p(s)[e(s, x) - T(x)]^2. \tag{7.1}$$

$\bar{S}$ being the subset of S consisting of all those samples for which $p(s) > 0$. According to the definition in Part II of this paper, the estimate $e(s, x)$ is admissible in the entire class of all estimates if there exists no other estimate $e_1(s, x)$ such that, for all $x \varepsilon R_N$

$$V(e_1, d) \leqq V(e, d) \tag{7.2}$$

the strict inequality in (7.2) holding for at least one $x \varepsilon R_N$.

However the experimental procedure which determines the sampling design d is generally, subject to certain limitations, under the statistician's control. If therefore there exists another design d_1, such that

$$V(e_1, d_1) \leqq V(e, d) \tag{7.3}$$

for all $x \varepsilon R_N$, the strict inequality holding for at least one $x \varepsilon R_N$, then we should use the design d_1, in conjunction with the estimate $e_1(s, x)$ in preference to the pair $(e(s, x), d)$. In practice, the class of available alternative designs d_1 is limited by the considerations of cost or time. In a case in which no such limitations existed, obviously the only admissible d is that which assigns probability 1 to the sample s_0 consisting of the whole population, so that taking

$$e(s_0, x) = \sum_{i=1}^{N} x_i = T(x);$$

$$V(e, d) \equiv 0 \qquad \text{for all} \quad x \varepsilon R_N.$$

In practice, through considerations of cost and time, the class of alternative designs d_1 is limited by one of the following conditions: viz (a) the average (i.e. expected) sample size does not exceed a certain limit or (b) the average cost of sampling does not exceed a certain limit. Condition (b) is obviously equivalent to (a) except when the cost of each observation depends on the unit observed. In practical situations, condition (a) is perhaps more commonly met with than (b).

Adopting condition (a) we therefore define 'uniform admissibility' as follows:

DEFINITION 7.A. An estimate $e_0(s, x)$ and a sampling design d_0 are uniformly admissible for the population total $T(x)$, if there does not exist any other estimate $e_1(s, x)$ and sampling design d_1 such that

(7.4) (i) expected sample size for $d_1 \leqq$ expected sample size for d_0 ;

(ii) $$V(e_1, d_1) \leqq V(e_0, d_0),$$

where V denotes the mean squared error as defined in (7.1), and the strict inequality holds in (7.4) either in (i) or for at least one $x \varepsilon R_N$ in (ii). Clearly this yields a stronger definition of admissibility as the estimate $e_0(s, x)$ forms a subclass of the class of admissible estimates for the design d_0 and similarly the design d_0 forms a sublcass of the class of designs which are admissible for fixed $e_0(s, x)$ (with an obvious definition for the admissibility of sampling designs for a fixed estimate $e(s, x)$).

We shall now prove the following theorem:

THEOREM 7.A. *The estimate* =

$$e^*(s, x) = [N/n(s)] \textstyle\sum_{i\varepsilon s} x_i \tag{7.5}$$

where $n(s)$ *is the sample size,* i.e. *the number of distinct units in the sample* s, *and a sampling design* d^* *of fixed sample size* (i.e. $p(s) = 0$ *unless* $n(s) =$ *some fixed number* m) *are uniformly admissible for the population total* $T(x)$ *in the sense of Definition* 7.A.

PROOF. If the theorem is not true, then there exists a sampling design d_1 and an estimate $e_1(s, x)$ such that

$$\text{expected sample size for } d_1 \leqq m \tag{7.6}$$

and

$$V(e_1, d_1) \leqq V(e^*, d^*), \tag{7.7}$$

where the strict inequality holds either in (7.6) or for at least one $x \varepsilon R_N$ in (7.7).

For the sampling design d_1 let

$$p_1(s) = \text{probability of sample } s,$$

$$n_1(s) = \text{size of sample } s$$

and $\bar{S}_1$: the subset of S all those sample s for which $p_1(s) > 0$ and let the corresponding terms for the sampling design d^* be $p^*(s)$, $n^*(s)$, and $\bar{S}^*$, respectively. Thus since by the theorem, d^* is a fixed sample size design

$$n^*(s) = m \qquad \text{for all} \quad s \varepsilon \bar{S}^* \tag{7.8}$$

and from (7.4) and (7.8)

$$\textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)n_1(s) \leqq \sum_{s\varepsilon\bar{S}^*} p^*(s)n^*(s) = m \tag{7.9}$$

and

(7.10) $\sum_{s\varepsilon\bar{S}_1} p_1(s)[e_1(s, x) - T(x)]^2 \leqq \sum_{s\varepsilon\bar{S}^*} p^*(s)[e^*(s, x) - T(x)]^2$

where the strict inequality holds either in (7.9) or for at least one $x \varepsilon R_N$ in (7.10).

We now take the expectations of both sides of (7.10) wrt a prior distribution on R_n, under which all the x_i ($i = 1, 2, \cdots, N$) are distributed independently and identically with common mean θ and variance σ^2 and get

(7.11) $\sum_{s\varepsilon\bar{S}_1} p_1(s)E[e_1(s, x) - T(x)]^2 \leqq \sum_{s\varepsilon\bar{S}^*} p^*(s)E[e^*(s, x) - T(x)]^2.$

Now in the left hand side of (7.11), putting

(7.12) $$g_1(s, x) = [N - n_1(s)]^{-1}[e_1(s, x) - \textstyle\sum_{i\varepsilon s} x_i];$$

(7.13) $$\begin{aligned} E[e_1(s, x) - \Gamma(x)]^2 &= E[(N - n_1(s))(g_1(s, x) - \theta) + \textstyle\sum_{i\not\varepsilon s} (x_i - \theta)]^2 \\ &= (N - n_1(s))^2E[g_1(s, x) - \theta]^2 + (N - n_1(s))\sigma^2, \end{aligned}$$

the cross terms vanishing due to the independence of the distribution of the x_i. Similarly in the right hand side of the (7.11) putting for any sample s

(7.14) $$\bar{x}_s = [n(s)]^{-1} \textstyle\sum_{i\varepsilon s} x_i$$

and using (7.8), we have

(7.15) $E[e^*(s, x) - T(x)]^2 = (N - m)^2E(\bar{x}_s - \theta)^2 + (N - m)\sigma^2.$

Now substituting (7.13) and (7.15) in (7.11) and using the relation

$$\textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s) = \sum_{s\varepsilon\bar{S}^*} p^*(s) = 1$$

and cancelling out the common term, (7.11) becomes

(7.16) $$\begin{aligned} \textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)(N - n_1(s))^2E(g_1(s, x) - \theta)^2 - \sigma^2 \sum_{s\varepsilon\bar{S}_1} p_1(s)n_1(s) \\ \leqq \textstyle\sum_{s\varepsilon\bar{S}^*} p^*(s)(N - m)^2E(\bar{x}_s - \theta)^2 - m\sigma^2. \end{aligned}$$

Now putting $g_1(s, x) = \bar{x}_s + h_1(s, x)$, and noting that if the sample size is $n(s)$, $E(\bar{x}_s - \theta)^2 = \sigma^2/n(s)$, we have from (7.16), after cancelling out the common term

(7.17) $$\begin{aligned} &\textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)(N - n_1(s))^2E(h_1^2(s, x)) \\ &\quad + 2 \textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)(N - n_1(s))E[h_1(s, x)(\bar{x}_s - \theta)] \\ &\quad\quad + \sigma^2N^2 \textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)/n_1(s) \leqq \sigma^2N^2/m. \end{aligned}$$

Now it follows from (7.9) that

(7.18) $$\textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)/n_1(s) \geqq m^{-1}$$

where the sign of equality holds if, and only if $n_1(s) = m$ for all $s \varepsilon \bar{S}_1$. Combining (7.17) and (7.18), we have

(7.19) $$\begin{aligned} &\textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)(N - n_1(s))^2E[h_1^2(s, x)] \\ &\quad + 2 \textstyle\sum_{s\varepsilon\bar{S}_1} p_1(s)[N - n_1(s)E[h_1(s, x)(\bar{x}_s - \theta)] \leqq 0. \end{aligned}$$

Now it is easily seen that (7.19) is equivalent to the inequality contained in clause (d), in the Lemma 1 of Part II of this paper and from the result proved therein, it follows that for all $s \varepsilon \bar{S}_1$

$$h_1(s, x) = 0 \tag{7.20}$$

so that $g_1(s, x) = \bar{x}_s$ and by (7.5) and (7.12),

$$e_1(s, x) = e^*(s, x). \tag{7.21}$$

Because of (7.20), the first two terms in the left hand side of (7.17) vanish, and hence the sign of equality must hold in both (7.17) and (7.18) so that d_1 is also a sampling design of fixed size m. Thus we must have in (7.7)

$$e_1(s, x) = e^*(s, x) \quad \text{so that} \quad V(e_1, d_1) = V(e^*, d_1) \tag{7.21.a}$$

where d_1 must be a sampling design of fixed sample size m. We shall next show that the sign of strict inequality in (7.7) cannot hold.

Let the inclusion probabilities for the units i ($i = 1, 2, \cdots, N$) and for the pair of units i, j ($i, j = 1, 2, \cdots, N$) for d_1 and d^* be given by

$$\Pi_{1i} = \sum_{s \ni i} p_1(s), \qquad \Pi_i^* = \sum_{s \ni i} p^*(s),$$
$$\Pi_{1ij} = \sum_{s \ni i,j} p_1(s), \qquad \Pi_{ij}^* = \sum_{s \ni i,j} p^*(s_i).$$

It is then easily found that

$$\begin{aligned} V(e^*, d_1) = {} & \sum_{i=1}^N x_i^2[\Pi_{1j}(N^2/m^2 - 2N/m) + 1] \\ & + 2 \sum_{i<j} x_i x_j [\Pi_{1ij} N^2/m^2 - (N/m)(\Pi_{1i} + \Pi_{ij}) + 1] \leqq 0 \end{aligned} \tag{7.22}$$

with a corresponding expression for $V(e^*, d^*)$. Now (7.7) clearly implies that the coeft of x_i^2 for each i, $i = 1, 2, \cdots, N$, in the right hand side of (7.22) must be $\leqq$ the coeft of x_i^2 in the corresponding expression for $V(e^*, d^*)$ as otherwise $V(e^*, d_1)$ will exceed $V(e^*, d^*)$ if we put $x_i = 1$ and all $x_i = 0$, $j \neq i$, $j = 1, 2, \cdots, N$. Thus we have from (7.22)

$$\Pi_{1i} \leqq \Pi_i^*, \qquad i = 1, 2, \cdots, N \tag{7.23}$$

But by a well known result

$$\sum_{i=1}^N \Pi_{1i} = \text{the expected sample size for} \quad d_1 = m$$

and similarly

$$\sum_{i=1}^N \Pi_i^* = \text{expected sample size for} \quad d^* = m.$$

Hence (7.23) implies that

$$\Pi_{1i} = \Pi_i^*, \qquad i = 1, 2, \cdots, N. \tag{7.24}$$

Next consider the expression (7.22) at the point $x \varepsilon R_N$, at which only two particular co-ordinates x_i and x_j differ from zero, so that in the expressions for $V(e^*, d_1)$ and $V(e^*, d^*)$ all coefficients other than those of the terms x_i^2, x_j^2 and

$2x_ix_j$ vanish, since by (7.24) the coefficients of the terms x_i^2 and x_j^2 are equal, we get

$$(7.25)\qquad 2x_ix_j\Pi_{1ij} \leqq 2x_ix_j\Pi^*_{ij} .$$

Since (7.25) holds for both positive and negative values of the product x_ix_j, we must have

$$(7.26)\qquad \Pi_{1ij} = \Pi^*_{ij} .$$

Clearly (7.26) must hold for each pair i, j. It now follows from (7.21), (7.24), (7.26) and (7.22) that in (7.7) $V(e, d_1) = V(e^*, d^*)$ for all $x \varepsilon R_N$ and the strict inequality in (7.7) does not hold for any $x \varepsilon R_N$. As d_1 has been shown to be necessarily of fixed sample size m, the strict inequality does not hold in (7.6) either. The theorem is thus proved.

8. Minimaxity of the estimate $e^*(s, x)$. The estimate $e^*(s, x)$ has been considered in Part II of this paper and proved to be admissible in the entire class of estimates for the population total. Now it is seen that the proof in Part II holds also for any subset of R_N which is symmetrical in all the co-ordinates x_r, $r = 1, 2, \cdots, N$. As observed in S.3 of Part II, Aggarwal (1959) has already established, the minimaxity of the estimate $e^*(s, x)$ in the subset D_0 of R_N given by

$$(8.1)\qquad D_0 = [x: \textstyle\sum_{r=1}^N (x_r - \bar{x})^2 \leqq \text{const} = N\sigma_0^2 \quad \text{say}]$$

where $\bar{x} = N^{-1}T(x)$, when the sampling design is that of simple random sampling with fixed size m. But a stronger result is seen to follow immediately from our result in Part II. For consider the subset D_σ of R_N given by

$$(8.2)\qquad D_\sigma = [x: \textstyle\sum_{r=1}^N (x_r - \bar{x})^2 = N\sigma^2].$$

Now let the sampling design be of simple random sampling, but not necessarily of fixed size, i.e. for each size m, $1 \leqq m \leqq N$, the total probability P_m is distributed equally between all possible samples of size m. Then by the usual formula, for points $x \varepsilon D$ in (8.2), the mean squared error of the estimate

$$e^*(s, x) = \textstyle\sum_{m=1}^N P_m(N^2\sigma^2/m)[1 - (m - 1)/(N - 1)] = a \quad \text{constant}.$$

Hence the admissibility of $e^*(s, x)$ for $x \varepsilon D_\sigma$ implies that the mean squared error of any other estimate is either equal to that of $e^*(s, x)$ for all $x \varepsilon D_\sigma$ or exceeds the latter for at least one point $x \varepsilon D_\sigma$. The minimaxity of $e^*(s, x)$ for $x \varepsilon D_\sigma$ and hence for the set D_0 in (9.1) follows.

In fact the result can be further generalized and shown to hold for any set S which is symmetrical in all the co-ordinates. Let σ_0^2 be the supremum of the MSE (mean squarred error) of $e^*(s, x)$ for $x \varepsilon S$, i.e.

$$(8.3)\qquad \sigma_0^2 = \sup_{x\varepsilon S} \text{MSE of } e^*(s, x).$$

Then for any arbitrary number ϵ, such that $0 < \epsilon < \sigma_0^2$, there exist points $x \varepsilon S$, for which

$$(8.4)\qquad \text{MSE of } e^*(s, x) \geqq \sigma_0^2 - \epsilon.$$

Let $(\sigma')^2 \geqq \sigma_0^2 - \epsilon$ be a value attained by the MSE of $e^*(s, x)$ for $x \varepsilon S$. Then the intersection of the set S and $D_{\sigma'}$, $D_{\sigma'}$ being the set as defined in (8.2) on substituting σ' for σ, is non-empty. Hence as the set $SD_{\sigma'}$ is symmetrical in all co-ordinates, $e^*(s, x)$ is admissible on it and hence for any other estimate $e_1(s, x)$

$$\max_{x\varepsilon D_{\sigma'}S} \text{MSE} \quad \text{of} \quad e_1(s, x) \geqq (\sigma')^2 \tag{8.5}$$

and hence $\sup_{x\varepsilon S}$ MSE of $e_1(s, x) \geqq (\sigma')^2 \geqq \sigma_0^2 - \epsilon$. As ϵ can be taken arbitrarily small it follows that

$$\sup_{x\varepsilon S} \text{MSE} \quad \text{of} \quad e_1(s, x) \geqq \sigma_0^2. \tag{8.6}$$

Our result is thus a generalization of that of Aggerwal (1959). Moreover Aggerwal's result was proved subject to the restriction of measurability of the estimate, while our result holds without this restriction.

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

REFERENCES

[1] AGGARWAL, O. P. (1959). Bayes and minimax procedures in sampling from finite and infinite populations—I. *Ann. Math. Statist.* **30** 206–218.

[2] DURBIN, J. (1953). Some results in sampling theory when units are selected with unequal probabilities. *J. Roy. Statist. Soc. Ser. B.* **15** 262–269.

[3] HAJEK, J. (1959). Optimum strategy and other problems in probability sampling. *Cosopis. Pest. Mat.* **84** 387–423.

[4] HODGES, J. L., JR., and LEHMANN, E. L. (1951). Some applications of the Cramer-Rao inequality. *Proceeding of the Second Berkeley Symp. Math. Statist. Prob.* 13–22. Univ. of California Press.

[5] JOSHI, V. M. (1965). Admissibility and Bayes estimation in sampling finite populations—II and III. *Ann. Math. Statist.* **36** 1723–1742.

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

[7] LOÈVE, MICHEL (1960). *Probability Theory* (2nd Edition-Application 2°). Wiley, New York. 126.

[8] ROGOSINSKI, WERNER W. (1952). *Volume and Integral.*

[9] WOLFOWITZ, J. (1947). The efficiency of sequential estimates and Wald's equation for sequential processes. *Ann. Math. Statist.* **18** 215–230.