ADMISSIBILITY AND BAYES ESTIMATION IN SAMPLING
FINITE POPULATIONS. II

By V. M. Joshi

University of North Carolina

1. Introduction. In Part I of this paper the admissibility was investigated primarily for the class of unbiased estimates of the population total. In particular the Horvitz-Thomson estimate was shown to be admissible in the class of all unbiased estimates, (cf. Theorem 4.1 of Part I). In the following, the investigation is extended by removing the restriction of unbiasedness, with the corresponding modification of the definition of admissibility: Now some other estimate is shown to remain admissible for all sampling designs. The result appears to have implications concerning the basic logic of sampling with varying probabilities. These however are not discussed here.

2. Notation. The notation used here is the same as that formulated in the Section 2 of the Part I of this paper and is not restated here. The definitions and preliminaries, as given in that section, also apply in the following discussion. In addition for convenience of discussion, here we assume that the units u of the population U are numbered, that is \( U = (u_1, \ldots, u_N) \), \( N \) being the total number of units \( u \) in \( U \). As a result a sample \( s \) (Definition 2.2, Part I) can now be specified by the set of integers namely the serial numbers of the units \( u \in s \). Thus for \( u \in s \) now we write \( r \in s \). Further, the variate value \( x(u_r) \) associated with the unit \( u_r \) would be denoted simply by \( x_r \), \( r = 1, \ldots, N \). And we have \( x = (x_1, \ldots, x_N) \), a point in Euclidean \( N \)-space \( R_N \). Now the problem is to find an estimate (Definition 2.6, Part I), of the population total

\[
T(x) = \sum_{r=1}^{N} x_r
\]

by observing those \( x_r \) for which \( r \in s \), the sample \( s \) being drawn according to a given sampling design (Definition 2.3, Part I). We extend the Definition 2.8, in Part I, of an admissible estimate by removing the restriction of unbiasedness as follows:

Definition. Given a sampling design \( d = (S, p) \), an estimate \( e(s, x) \) is said to be admissible for \( T \) in (1), if and only if there does not exist any other estimate \( e'(s, x) \) such that

\[
\sum_{s \in S} p(s)(e'(s, x) - T(x))^2 \leq \sum_{s \in S} p(s)e(s, x) - T(x))^2
\]

for all \( x \in R_N \), strict inequality holding true for at least one \( x \).

3. Admissibility of an estimate. We now prove the following

Theorem. The estimate \( e^*(s, x) \) given by

\[
e^*(s, x) = \left( \frac{N}{n(s)} \right) \sum_{r \in s} x_r
\]

Received 25 March 1964; revised 28 May 1965.

On leave from Maharashtra Government, Bombay.
where \( n(s) \) is the sample size (Definition 2.4, Part I), is admissible for \( T \) according to the Definition in the preceding section, for any sampling design.

Remark. \( e^*(s, x) \) can also be shown to be admissible on any subset of \( R_N \) given by \( x: c_1 \leq x_r \leq c_2, r = 1, \ldots, N, c_1, c_2 \) being some arbitrary constants with a slight obvious modification of the proof below.

Proof. If \( e^* \) in (3), is not admissible, then by (2) there exists an estimate \( e'(s, x) \) such that, for all \( x \in R_N \),

\[
(4) \quad \sum_{s \in S} p(s) (e'(s, x) - T(x))^2 \leq \sum_{s \in S} p(s) (e^*(s, x) - T(x))^2.
\]

We put

\[
(5) \quad g(s, x) = (N - n(s))^{-1} e'(s, x) - \sum_{r \in S} x_r,
\]

\[
(6) \quad g^*(s, x) = (N - n(s))^{-1} e^*(s, x) - \sum_{r \in S} x_r,
\]

\( n(s) \) being the sample size (Definition 2.4, Part I) of \( s \). Now assuming \( n(s) = N \rightarrow p(s) = 0 \), and putting for such \( s \), \( g = g^* = c \) in (5), we have from (4)

\[
(6) \quad \sum_{s \in S} p(s) [(N - n(s)) g(s, x) - \sum_{r \in S} x_r]^2
\]

\[
\leq \sum_{s \in S} p(s) [(N - n(s)) g^*(s, x) - \sum_{r \in S} x_r]^2.
\]

(Even without this assumption, the proof needs only a slight modification. For, obviously it is enough to consider in (4) estimates \( e' \) such that \( e' = T \), for sample \( s \) for which \( n(s) = N \).) Now taking the expectations of both sides of (6) wrt a probability distribution of \( R_N \) such that \( x_1, \ldots, x_N \) are independently and identically distributed, with a common finite discrete frequency function \( w \), common mean \( \theta(w) \) and common variance \( \sigma^2(w) \), we have

\[
(7) \quad \sum_{s \in S} p(s) (N - n(s))^2 E_w [(g(s, x) - \theta(w)) + (\theta(w)
\]

\[
- (N - n(s))^{-1} \sum_{r \in S} x_r]^2 \leq \sum_{s \in S} p(s) (N - n(s))^2 E_w
\]

\[
\cdot [(g^*(s, x) - \theta(w)) + (\theta(w) - (N - n(s))^{-1} \sum_{r \in S} x_r)]^2.
\]

The existence of \( E_w \) in (7) follows from the finite discreteness of the frequency function \( w \). Now noting that the expectations of the product terms on both sides of (7) vanish due to the independence of \( x_1, \ldots, x_N \) and cancelling out the common term \( \sum_{s \in S} p(s)(N - n(s))^2 \sigma^2(w) \) on both sides of (7), we get

\[
(8) \quad \sum_{s \in S} p(s) (N - n(s))^2 E_w (g(s, x) - \theta(w))^2
\]

\[
\leq \sum_{s \in S} p(s)(N - n(s))^2 E_w (g^*(s, x) - \theta(w))^2.
\]

Since \( x_r, r = 1, \ldots, N \) are distributed independently and identically we replace in \( h(s, x) \) and \( h^*(s, x) \) in (8) the variates \( x_r, r \in s \), in some order by \( x_1, x_2, \ldots, x_m \) respectively, and let

\[
(9) \quad h(s, x) \text{ and } h^*(s, x) \text{ denote the resulting values of } g(s, x) \text{ and } g^*(s, x), \text{ respectively.}
\]
Next putting in (7),
\[ \sum_{s \in S_m} p(s) h(s, x) = P_m \phi_m(x) \]
where \( S_m \) is the set of all samples \( s \) with fixed size \( m \), i.e. \( n(s) = m \) and \( P_m = \sum_{s \in S_m} p(s) \), we have
\[ \sum_{s \in S_m} p(s) (N - n(s))^2 E_w(h(s, x) - \theta(w))^2 \]
where \( S_m \) is the set of all samples \( s \) with fixed size \( m \), i.e. \( \sum_{s \in S_m} p(s) = m \) and \( P_m = \sum_{s \in S_m} p(s) \). We then have
\[ \sum_{s \in S_m} p(s)(N - n(s))^2 E_w(h(s, x) - \phi_m(x))^2 \]
where \( S_m \) is the set of all samples \( s \) with fixed size \( m \), i.e. \( \sum_{s \in S_m} p(s) = m \) and \( P_m = \sum_{s \in S_m} p(s) \).

Now if in (10) \( h(s, x) \) is replaced by \( h^*(s, x) \) in (5) and \( \phi_m(x) \) by \( \phi^*_m(x) \), then from (3), we get
\[ h^*(s, x) = \phi^*_m(x) = \sum_{r=1}^{n(s)} x_r / n(s). \]
Hence from (11) and (12)
\[ \sum_{s \in S_m} p(s)(N - n(s))^2 E_w(h^*(s, x) - \theta(w))^2 \]
\[ = \sum_{m=1}^{N} P_m (N - m)^2 (\phi^*_m(x) - \theta(w))^2. \]
And further from (8), (11) and (13) we get
\[ \sum_{m=1}^{N} (N - m)^2 \sum_{s \in S_m} p(s) E_w(h(s, x) - \phi_m(x))^2 \]
\[ + \sum_{m=1}^{N} (N - m)^2 P_m E_w(\phi_m(x) - \theta(w))^2 \]
\[ \leq \sum_{m=1}^{N} (N - m)^2 P_m E_w(\phi^*_m(x) - \theta(w))^2. \]
That is
\[ \sum_{m=1}^{N} (N - m)^2 P_m E_w(\phi_m(x) - \theta(w))^2 \]
\[ \leq \sum_{m=1}^{N} (N - m)^2 P_m E_w(\phi^*_m(x) - \theta(w))^2. \]
Now from (15) and Lemma 1 in the next section we get if \( P_m \not= 0 \),
\[ \phi_m(x) = \phi^*_m(x) \]
for all \( x \in R_N \). Further, substituting (16) in (14) we have
\[ E_w(h(s, x) - \phi^*_m(x))^2 = 0 \]
for all samples \( s \) having \( p(s) \not= 0 \). Next from (17) and Lemma 2, in the next section, we have
\[ h(s, x) = \phi^*_m(x) \]
for all \( s \) having \( p(s) \not= 0 \) and all \( x \). Further from (5), (12), (18) and (19) follows the result
\[ e'(s, x) = e^*(s, x). \]
Now (4) and (19) imply the Theorem stated at the beginning of this section.

It is interesting to note that using a result due to Hodges and Lehmann (1951)
establishing the admissibility of sample mean, wrt squared error as loss, for the
mean of a normal population with unit variance, we can from (15) straightaway
deduce, that a.e. in \( R_m \),

\[
\phi_m(x) = \phi_m^*(x)
\]

for a fixed sample size design (i.e. \( p(s) = 0 \) if \( n(s) \neq m \)). Note here we have not
used Lemma 1. Apart from the restriction of fixed sample size design in (20), it
is important that \( \phi_m(x) = \phi_m^*(x) \) in (20) is established for almost all points in
\( R_m \); while what we need for establishing our ultimate result is \( \phi_m(x) = \phi_m^*(x) \)
for all points in \( R_m \), which is achieved in (16) with the help of Lemma 1.

It is also worth while to note that Aggarwal (1959) has already investigated
the minimaxity of the estimate \( e^*(s, x) \) in (3), on a certain subset of \( R_N \). However
he restricts himself to simple random sampling without replacement with
fixed number of draws. In contrast, we establish the admissibility of the
estimate \( e^* \) for any sampling design (Definition 2.3, Part I) what so ever. Further
the subset of \( R_N \) considered by Aggarwal is given by \( x = (x_1, \ldots, x_N) : \sum_{r=1}^N (x_r - T(x)/N)^2 \leq \text{const.} \) while our Remark following the Theorem in this
section establishes the admissibility of \( e^*(s, x) \) on a practically much more
realistic subset of \( R_N \) as explained in Section 3 of Part I of this paper.

4. Lemmas. Now we would prove the lemmas referred to in the last section.

**Lemma 1.** If

(a) \( x_1, x_2, \ldots, x_N \) are independently and identically distributed real random
variates,

(b) for every \( m = 1, \ldots, N, \phi_m(x) \) is a real function of \( x_1, x_2, \ldots, x_m \),

(c) for every \( m = 1, \ldots, N, \bar{x}_m = (1/m) \sum_{i=1}^m x_i \),

(d) for every common finite discrete frequency function \( w \) of \( x_1, \ldots, x_N \),

\[
\sum_{m=1}^N A_m E_w(\phi_m(x) - \theta(w))^2 \leq \sum_{m=1}^N A_m E_w(\bar{x}_m - \theta(w))^2,
\]

\( E_w \) denoting the expectation, \( \theta(w) \) the common mean of \( x_1, \ldots, x_N \) and \( A_m \),
\( m = 1, \ldots, N \) being arbitrary real constants, then for every \( x = (x_1, x_2, \ldots, x_N) \in R_N \), \( \phi_m(x) = \bar{x}_m \) for all \( m = 1, \ldots, N \) for which \( A_m \neq 0 \).

**Proof.** Let \( B_k \subset R_N \) be such that if \( x = (x_1, \ldots, x_r, \ldots, x_N) \in B_k \) then
\( x_r, r = 1, \ldots, N \) contain \( k \) or less distinct values. Now by the condition (d)
of the Lemma 1, considering the discrete frequency function \( w \) which is zero
every where except at one point, we have, for all \( x \in B_1 \),

(1*) \[ \phi_m(x) = \bar{x}_m \] for all \( m = 1, \ldots, N \) such that \( A_m \neq 0 \).

Further in the next paragraph, we prove that if (1*) holds for \( x \in B_{k-1} \) then it also
holds for all \( x \in B_k \), which would mean (1*) holds for all \( x \in B_N = R_N \), proving
the Lemma 1.

Let the common frequency function of \( x_1, \ldots, x_N \), referred to in the condition (d) of the Lemma 1, be zero except at \( k \) specified distinct values namely,
where $w(t_i) = p_t$, $p_t > 0$, $i = 1, \ldots, k$ and $\sum_{i=1}^k p_t = 1$. This frequency function clearly gives positive probability only to those points $x = (x_1, \ldots, x_r, \ldots, x_N)$ for which $x_r$, $r = 1, \ldots, N$ is one of the values $t_1, \ldots, t_k$. Let these points $x$ constitute the set $B_k(t_1, \ldots, t_k)$. Then $B_k(t_1, \ldots, t_k) \subseteq B_h$ defined in the beginning of this proof.

Throughout the remainder of the proof, summations over all $x \in B_k(t_1, \ldots, t_k)$, $x(m) \in D_{mk}(t_1, \ldots, t_k)$ and $x(m) \in D'_{mk}(t_1, \ldots, t_k)$ will be indicated by $\sum_{B_k}$, $\sum_{D_{mk}}$ and $\sum_{D'_{mk}}$, respectively.

Now writing

$$\phi_m(x) = x_m + h_m(x),$$

we have from (d)

$$\sum_{m=1}^N A_m^2 \sum_{B_k} h_m(x)(x_m - \theta) \prod_{i=1}^k p_i^{g(t_i, x)} \leq 0,$$

$g(t_i, x)$ denoting for each $x = (x_1, \ldots, x_r, \ldots, x_N)$ the total number of those $x_r$, $r = 1, \ldots, N$, which are equal to $t_i$. Note, for all $x \in B_k(t_1, \ldots, t_k)$, $\sum_{i=1}^k g(t_i, x) = N$, $g(t_i, x) \geq 0$ and

$$\theta = \sum_{i=1}^N p_d t_i.$$

Now let $D_{mk}(t_1, \ldots, t_k) \subseteq R_m$ the $m$-space of the points $x(m) = (x_1, \ldots, x_m)$, the first $m$ coordinates of $x = (x_1, \ldots, x_N)$, such that

$$x(m) \in D_{mk}(t_1, \ldots, t_k) \text{ if and only if } x \in B_k(t_1, \ldots, t_k).$$

Since $h_m(x)$ and $x_m$ are defined on $R_m$, by summing in (3*) for all $x \in B_k(t_1, \ldots, t_k)$ with a common $x(m)$, we have,

$$\sum_{m=1}^N A_m^2 \sum_{D_{mk}} h_m(x)(x_m - \theta) \prod_{i=1}^k p_i^{g(t_i, x(m))} \leq 0,$$

where $g(t_i, x(m))$ is the total number of co-ordinates in $x(m) = (x_1, \ldots, x_m)$ which are equal to $t_i$, $i = 1, \ldots, k$. Note that for every $x(m) \in D_{mk}(t_1, \ldots, t_k)$, $g(t_i, x(m)) \geq 0$, $i = 1, \ldots, k$, $\sum_{i=1}^k g(t_i, x(m)) = m$, and

$$(1/m) \sum_{i=1}^k t_i g(t_i, x(m)) = x_m.$$

Now in (6*) let

$$D_{mk}(t_1, \ldots, t_k) = D'_{mk}(t_1, \ldots, t_k) + D^2_{mk}(t_1, \ldots, t_k),$$

where $x(m) = (x_1, \ldots, x_m) \in D'_{mk}(t_1, \ldots, t_k)$ if and only if $x_1, \ldots, x_m$ contain all the distinct values $t_1, \ldots, t_k$. Now we assume that (1*) holds for $x \in D_{mk-1}$. Since this assumption obviously means $A_m \neq 0 \Rightarrow h_m(x) = 0$ if the coordinates of $x(m)$ contain less than $k$ distinct values, we have for $m = 1, \ldots, N$,

$$(9*) \text{ if } A_m \neq 0 \text{ in } (8*) \text{ for all } x(m) \in D^2_{mk}(t_1, \ldots, t_k), \text{ } h_m(x) = 0.$$

From (6*) and (9*)

$$\sum_{m=1}^N A_m^2 \sum_{D^2_{mk}} h_m(x)(x_m - \theta) \prod_{i=1}^k p_i^{g(t_i, x(m))} \leq 0.$$
We note that in \((10^*)\),
\[
g(t_i, x(m)) \geq 1, \quad i = 1, \ldots, k.
\]
Next we substitute \((4^*)\) and \((7^*)\) in the left hand side of \((10^*)\) and multiply it by \(1/\Pi_{i=1}^{k} p_i\). The resulting expression (note here \((11^*)\)) is further integrated over the domain
\[
Q = [p_1, \ldots, p_k: p_i > 0, i = 1, \ldots, k \text{ and } \sum_{i=1}^{k} p_i = 1].
\]
We then have
\[
\sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} \int_{Q} h_m(x)(\bar{x}_m - \theta) \prod_{i=1}^{k} p_i^{g(t_i, x(m)) - 1} \Pi_{i=1}^{k-1} dp_i
\]
\[
= \sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} h_m(x) \int_{Q} (\sum_{j=1}^{k} (g(t_j, x(m))/m - p_j)t_j)
\]
\[
\cdot \prod_{i=1}^{k-1} p_i^{g(t_i, x(m)) - 1} \Pi_{i=1}^{k-1} dp_i
\]
\[
= 0,
\]
as for every \(j\),
\[
\int_{Q} t_j(g(t_j, x(m))/m - p_j) \prod_{i=1}^{k} p_i^{g(t_i, x(m)) - 1} \Pi_{i=1}^{k-1} dp_i = 0.
\]

[Note that:
\[
\int_{Q} \prod_{i=1}^{k} p_i^{n_i - 1} \Pi_{i=1}^{k-1} dp_i = [\Gamma(\sum_{i=1}^{k} n_i)]^{-1} \prod_{i=1}^{k} \Gamma(n_i) \quad \text{for } n_i \geq 1, i = 1, \ldots, k
\]

Now because of \((10^*)\) the integrand in \((12^*)\) \(\leq 0\) and is also continuous in \(p = (p_1, \ldots, p_k)\) for all \(p \in Q\). Therefore from \((12^*)\), we have
\[
\sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} h_m(x)(\bar{x}_m - \theta) \prod_{i=1}^{k} p_i^{g(t_i, x(m)) - 1} = 0
\]
for all \(p \in Q\). Next the condition (d) of the Lemma also gives in place of \((3^*)\), the stronger relation
\[
\sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} [h_m^2(x) + 2h_m(x)(\bar{x}_m - \theta)] \prod_{i=1}^{k} p_i^{g(t_i, x)} \leq 0.
\]
Then proceeding exactly as from \((3^*)\) to \((10^*)\) and lastly dividing by \(\prod_{i=1}^{k} p_i\), from \((14^*)\), we have for all \(p \in Q\),
\[
\sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} [h_m^2(x) + 2h_m(x)(\bar{x}_m - \theta)] \prod_{i=1}^{k} p_i^{g(t_i, x)} \leq 0.
\]
Further from \((13^*)\) and \((15^*)\) we get
\[
\sum_{m=1}^{N} A_m^2 \sum_{b_{nk}} h_m(x) \prod_{i=1}^{k} p_i^{g(t_i, x)} \leq 0
\]
for all \(p \in Q\). Next considering the inequality \((16^*)\) for a point \(p = (p_1, \ldots, p_k) \in Q\), we have
\[
A_m \neq 0 \Rightarrow h_m(x) = 0 \quad \text{for all } x(m) \in D_{mk}^t (t_1, \ldots, t_k).
\]
Thus from \((8^*)\), \((9^*)\) and \((17^*)\) we have, for \(m = 1, \ldots, N\),
\[
A_m \neq 0 \Rightarrow h_m(x) = 0 \quad \text{for all } x(m) \in D_{mk}(t_1, \ldots, t_k).
\]
But since \( h_m(x) \) is a function of \( x_1, \cdots, x_m \) we have from (5*), (18*)

\[
(19*) \quad A_m \neq 0 \Rightarrow h_m(x) = 0 \quad \text{for all } x \in B_k(t_1, \cdots, t_k).
\]

Further since the set \( B_k \) as defined in the beginning of this proof satisfies \( B_k = \bigcup_{t_1, \cdots, t_k} B_k(t_1, \cdots, t_k) \), we have from (19*), for \( m = 1, \cdots, N, A_m \neq 0 \Rightarrow h_m(x) = 0 \text{ for all } x \in B_k \), which along with (2*) means that, for \( m = 1, \cdots, N, \)

\[
(20*) \quad A_m \neq 0 \Rightarrow \phi_m(x) = \bar{x}_m \quad \text{for all } x \in B_k.
\]

Thus as stated in the first paragraph of this proof, the Lemma 1 is proved by induction.

**Lemma 2.** If

(a) \( x_1, \cdots, x_m \) are independently and identically distributed real random variates,

(b) \( G(x) \) and \( H(x) \) be real functions of \( x = (x_1, \cdots, x_m) \in R_m \),

(c) for every common discrete frequency function \( w \) of \( x_1, \cdots, x_m \),

\[ E_w(g(x) - H(x))^2 = 0, \]

then \( G(x) = H(x) \) for all \( x = (x_1, \cdots, x_m) \in R_m \).

**Proof.** Let the common frequency function \( w \) in the condition (c) of this Lemma be zero, except at \( m \) specified values, namely \( w(t_i) = p_i, p_i > 0, i = 1, \cdots, m \) and \( \sum_{i=1}^{m} p_i = 1 \). This frequency function clearly gives positive probability say \( P(x) \) only to those points \( x = (x_1, \cdots, x_r, \cdots, x_m) \) for which \( x_r, r = 1, \cdots, m \) is one of the values \( t_1, \cdots, t_m \). Let these points \( x \), constitute the set \( B(t_1, \cdots, t_m) \). So that in condition (c) of this Lemma,

\[ E_w(G(x) - H(x))^2 = \sum_{x \in B(t_1, \cdots, t_k)} P(x)(G(x) - H(x))^2 = 0, \]

which implies \( G(x) = H(x) \) for all \( x \in B(t_1, \cdots, t_m) \) and as \( t_1, \cdots, t_m \) are arbitrary, the result \( G(x) = H(x) \) for all \( x \in R_m \) follows.

**Acknowledgment.** The author is indebted to Professor V. P. Godambe for guidance in this work, to Dr. J. Hájek for many helpful suggestions, to the referee for a thorough going scrutiny of this paper, and to Miss Patkar for doing the necessary typing very carefully.

**References**
