BAYES AND MINIMAX PROCEDURES IN SAMPLING FROM FINITE POPULATIONS

BY
OM PRAKASH AGGARWAL

TECHNICAL REPORT NO. 15

PREPARED UNDER CONTRACT N6onr-251 TASK ORDER III (NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

FEBRUARY 17, 1953
BAYES AND MINIMAX PROCEDURES IN SAMPLING
FROM FINITE POPULATIONS

BY
OM PRAKASH AGGARWAL

TECHNICAL REPORT NO. 15
FEBRUARY 17, 1953

PREPARED UNDER CONTRACT N6onr-251 TASK ORDER III
(NR-042-993)
FOR
OFFICE OF NAVAL RESEARCH

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I  Introduction</td>
<td>1</td>
</tr>
<tr>
<td>II Summary of Results</td>
<td>3</td>
</tr>
<tr>
<td>III Sampling from a Finite Population</td>
<td>6</td>
</tr>
<tr>
<td>IV Bayes and Minimax Estimates</td>
<td>6</td>
</tr>
<tr>
<td>V Bayes and Minimax Procedures for Estimating the Proportion Defective in a Lot of Finite Size</td>
<td>9</td>
</tr>
<tr>
<td>VI Bayes and Minimax Procedures for Estimating the Mean from a Stratified Population</td>
<td></td>
</tr>
<tr>
<td>A. Infinite Population</td>
<td>14</td>
</tr>
<tr>
<td>B. Finite Populations</td>
<td>20</td>
</tr>
<tr>
<td>VII Bayes and Minimax Procedures for Estimating the Mean from a Population with Two-Stage (Cluster) Sampling</td>
<td>27</td>
</tr>
<tr>
<td>A. Infinite Population</td>
<td>28</td>
</tr>
<tr>
<td>B. Finite Populations</td>
<td>35</td>
</tr>
<tr>
<td>C. Cluster Sampling</td>
<td>43</td>
</tr>
<tr>
<td>VIII Weighted Estimate as a Minimax Procedure</td>
<td></td>
</tr>
<tr>
<td>A. Infinite Population</td>
<td>44</td>
</tr>
<tr>
<td>B. Finite Populations</td>
<td>47</td>
</tr>
<tr>
<td>IX Ratio Estimate as a Minimax Procedure</td>
<td>50</td>
</tr>
<tr>
<td>A. Infinite Population</td>
<td>51</td>
</tr>
<tr>
<td>B. Finite Populations</td>
<td>54</td>
</tr>
<tr>
<td>X  Regression Estimate as a Minimax Procedure</td>
<td></td>
</tr>
<tr>
<td>A. Infinite Population</td>
<td>58</td>
</tr>
<tr>
<td>B. Finite Populations</td>
<td>61</td>
</tr>
<tr>
<td>XI Literature Cited</td>
<td>69</td>
</tr>
<tr>
<td>XII Acknowledgment</td>
<td>70</td>
</tr>
</tbody>
</table>
BAYES AND MINIMAX PROCEDURES IN SAMPLING
FROM FINITE POPULATIONS

By
Om Prakash Aggarwal

I. Introduction.

In a statistical decision situation involving two alternative hypotheses, the errors of the first and second kind are defined as the errors resulting from making decisions, viz., of rejecting an hypothesis when it is true and accepting it when it is false. In emphasizing the importance of these two types of errors in testing statistical hypotheses, Neyman and Pearson\(^1\) in 1933 unfolded the principle that a statistical procedure should be evaluated by its consequences in various circumstances. In 1939, Wald\(^2\) proposed the extension of this principle to all statistical problems and introduced the notions of weight and risk function and in later writing developed the general theory of statistical decision functions\(^3\).

The basic idea, emphasized by Wald, is that allocation of resources to gathering information must be balanced against the loss due to possible wrong decisions based on partial information. This idea, which has had such a profound influence on current statistical research, is interestingly enough, not new to statisticians in the field of sample surveys. In fact, it has been the main guide in both the theory of sampling from a finite population and the application of this theory, since the publication of


Neyman's fundamental paper in this field in 1934.\(^1\)

In the current practice of conducting sample surveys, the statisticians have adopted one of the following two procedures:\(^2\)

(i) To get an estimate of maximum precision for a given total cost of the survey, or
(ii) To get an estimate of given precision for a minimum total cost of the survey.

The allocation of the resources for a given survey is usually carried out, keeping in mind one or the other of the above two aims. It is obvious, however, that a more general approach would be to consider the losses resulting from the errors in the estimates and from the cost of sampling jointly, and to employ such sampling and estimation procedures as would minimize the total expected loss. This approach is the decision function approach and will be adopted in this paper. The primary aim was to investigate the assumptions in modern decision theory needed to yield the classical results in estimation and design. Such a study was considered desirable as a preliminary step toward further research in this field.

It should also be pointed out that attention is here confined only to non-sequential estimates, i.e., estimates based on a preassigned number of observations. There is no need to consider randomized estimates since the loss function to be considered is in each case convex and of finite


expectations; under those conditions Hodges and Lehman\(^1\) have shown that the class of non-randomized estimates is essentially complete. No attempt was made to investigate the admissibility of the procedures and estimates obtained but it is conjectured that they are all admissible.

II. Summary of Results.

The problem of sampling from a finite population has been considered in the following way. The finite population is regarded as the outcome of a fixed sample size experiment performed by nature or some conscious being, using some probability distribution unknown to the statistician. The loss function, however, does not depend upon the form of the probability distribution but only upon the outcome of this large experiment. Use has been made of the fact that under these circumstances, if the statistician is permitted to choose a sample of fixed size \(n\), the optimum strategy for him is to choose it by the method of simple random sampling.

In the case of the estimation of the proportion defective, say \(p\), in a lot of finite size, \(N\), it has been shown that the usual procedure of estimating \(p\) by the proportion found defective in a sample drawn at random from the lot is in fact a minimax procedure, while the least favorable distribution of nature is a uniform discrete distribution over all the \(N+1\) possible values of \(p\). The loss function in this case is assumed to be

\[
L(p,r) = (f-p)^2/p(1-p)
\]

and the constant risk obtained is \(\frac{1}{N} \frac{N-n}{N-1}\), where \(n\) is the size of the sample.

In the other cases of estimation considered, the least favorable distributions of nature do not exist. The minimax estimates are obtained by first obtaining a sequence of Bayes estimates and the corresponding Bayes risks, and then obtaining an estimate which has its risk equal to the limiting value of the Bayes risks. The Bayes and minimax estimates and sampling procedures have been obtained both for the infinite and the finite populations.

In the case of stratified sampling, the loss function is taken as

\[ L(u,f) = (f-u)^2 + \sum_{i=1}^{k} c_i n_i \]

where \( f \) is the estimate of \( u \), the population total, \( c_i, n_i \) are cost per sampling unit and the number of sampling units chosen from the \( i \)th stratum, while \( k \) is the number of strata. The minimax procedure has been found to choose \( n_i \) sampling units at random from the \( i \)th stratum, where

\[ n_i = \text{integer nearest to} \sqrt{\frac{N_i^2 \sigma_i^2}{c_i} + 1} \]

and to employ the usual estimate \( \bar{x}_i = \frac{1}{N_i} \sum_{i=1}^{k} N_i \bar{x}_i \) where \( N_i \) is the size of the \( i \)th stratum and \( \bar{x}_i \) the sample mean from that stratum. The minimax risk is found to be

\[ \sum_{i=1}^{k} [N_i^2 (\frac{1}{n_i} - \frac{1}{N_i}) \sigma_i^2 + c_i n_i]. \]

In the case of cluster sampling, which is considered as a special case of two stage sampling where sampling at the second stage consists of observing the entire cluster, the minimax risk, apart from sampling costs, is shown to be

\[ R = \sigma^2 \frac{N}{nM} [1 + (M-1) \rho] \]

where \( \sigma^2 \) is the variance of population sampled, \( M \) is the size of each cluster, and \( \rho \) the intra-class correlation.
In the consideration of the ratio method of estimation, it is shown that if a population can be classified on the basis of the values of an auxiliary characteristic \( y \), and if the unknown mean and the known variance of \( x \)'s for a given \( y \) are each proportional to \( y \), then a minimax procedure is to choose a random sample of size \( n_i \) from each class, \( n_i \) being proportional to \( N_i \), the total number of units in the \( i \)th class, and to employ the classical ratio method of estimation, viz.,

\[
f = \frac{\text{sum of } x \text{'s in the sample}}{\text{sum of } y \text{'s in the sample}} \times \frac{\text{sum of } y \text{'s in the sample}}{\text{sum of } y \text{'s in the population}}.
\]

The constant of proportionality for choosing \( n_i \) is given by

\[
n_i = \sqrt{\frac{\beta T_y}{C}}
\]

where \( \beta \) is the known constant of proportionality for the variances of \( x \) for given \( y \), \( T_y \) is the sum of all \( y \) values in the entire finite population and \( C \) is the cost of complete enumeration. Here also the loss function is taken to be the usual squared difference plus a constant cost per observation.

Finally it is shown that for any bivariate distribution with \( v \), the mean and \( \sigma_y^2 \), the variance of \( y \) known, \( \rho \), the correlation coefficient, and \( \sigma_x^2 \), the variance of \( x \) known, the minimax procedure for estimating \( u \), the unknown mean of \( x \) is to choose \( n \) pairs \((x, y)\) of observations, where

\[
n = \text{integer nearest to} \sqrt{\frac{\sigma_x^2 (1 - \rho^2)}{C} + \frac{1}{4}}
\]

and to employ the usual regression estimate

\[
\hat{x} = \frac{\rho \sigma_x^2}{\sigma_y^2} (y - v).
\]

Here \( C \) is the cost of sampling per pair of observations. The same result holds good in sampling from a finite population also. The loss function in this case has been taken as

\[
L(u, f) = (f - u)^2 + cn.
\]
III. Sampling from a finite population.

In terms of the concepts of decision theory, Blackwell and Girshick\(^1\) in their forthcoming book have formulated the problem of sampling from a finite population in the following manner: We are given a sample space and nature (or a conscious being) performs a fixed sample size experiment and obtains a value of a random variable, which is a point \(x = (x_1, x_2, \ldots, x_N)\) in a space \(X\), where each \(x_i\) is a real number. The statistician has to select one out of a class \(A\) of possible actions in complete or partial ignorance of \(x\) and \(\omega\), the particular probability distribution employed by nature to obtain \(x\). He (the statistician) incurs a loss \(L\) which is a bounded function of the action \(a \in A\) selected and the point \(x\), but usually not of the underlying \(\omega\). Moreover, for a given \(a\), \(L(x,a)\) is assumed to be constant for all permutations of the coordinates of \(x\). The statistician can obtain partial information on \(x\) by observing a fixed number of coordinates of \(x\), say \(n\). The problem is: if the cost of observing \(x_i\) is independent of \(i\), how should the statistician select the sample?

It is proved there that the principle of invariance together with the principle of sufficiency leads to a strategy which selects each set \(s\) of \(n\) distinct integers from 1 to \(N\) (without regard to order) with probability \(\frac{1}{\binom{N}{n}}\), which is the strategy of simple random sampling.

IV. Bayes and minimax estimates.

The estimation problem with a fixed sample size has the following structure. We are given a sample space \(X\), a space of probability distributions \(\Omega\) and a numerical valued function \(\theta\) defined on \(\Omega\) whose value the statistician wishes to estimate on the basis of the outcome of an experiment, say \(x \in X\).

\(^1\)David Blackwell and M. A. Girshick, *Theory of Games and Statistical Decisions*, John Wiley and Sons (in press) Chapter VIII.
A non-randomized decision function for the statistician, called an (non-randomized) estimate, is a numerical function \( \delta \) defined on \( X \), specifying for each \( x \in X \), the number \( a \in A \) which will be chosen to estimate \( \theta(\omega) \) when that \( x \) is observed. The space of actions \( A \) is here the real line. The loss function, defined on \( \Omega \times A \) is non-negative, and is the loss incurred when \( \theta(\omega) \) is estimated by \( \delta(x) \). The risk function \( R \) is defined by

\[
R(\omega, \delta) = \mathbb{E}_\omega L[\theta(\omega), \delta]
\]

The subscript \( \omega \) appended to the symbol \( \mathbb{E} \) for expectation indicates that \( \omega \) is to be regarded as fixed when expectation is taken.

If we assume that \( \omega \) is given by nature as a random variable with a probability distribution \( \lambda \), we define a Bayes estimate with respect to the a priori distribution \( \lambda \) as that estimate \( \delta \) which minimizes the average risk

\[
\int R(\omega, \delta) d\lambda(\omega).
\]

If the statistician knew \( \lambda \), he would choose this estimate as his best decision. But usually \( \lambda \) is unknown, and the statistician may decide to use a minimax strategy. We define an estimate \( \delta \) to be minimax if it minimizes

\[
\sup_{\omega \in \Omega} R(\omega, \delta).
\]

If nature has a distribution which is least favorable to the statistician, it is called a maximin strategy and is defined as the distribution \( \lambda \) such that it maximizes

\[
\inf_{\delta} \int R(\omega, \delta) d\lambda(\omega).
\]

The following theorem, \footnote{E. L. Lehmann, Mimeographed notes on the Theory of Estimation, University of California, Berkeley, Chapter IV.} restated for reference purposes, will give both the minimax estimate and the least favorable distribution, whenever the latter exists.

\textbf{Theorem 4.1.} If a Bayes estimate \( \delta_\lambda \) has constant (independent of \( \omega \)) risk \( R(\omega, \delta_\lambda) = r \), then \( \delta_\lambda \) is minimax, and \( \lambda \) is a least favorable distribution.
of nature.

There are many problems in which no least favorable distribution exists. In none of the problems considered here does a least favorable distribution exist, excepting in the problem of estimation of the proportion defective in a lot of finite size, discussed in the next section. For those problems where no least favorable distribution exists, we have applied the following theorem. It is also proved in the notes on the Theory of Estimation by Lehmann. 1/

Theorem 4.2. If \( \{\lambda_n\} \) be a sequence of a priori probability distributions, \( \{\delta_n\} \) the sequence of associated Bayes estimates, \( \{r_n\} \) the sequence of associated Bayes risks, and if \( r_n \to r \) as \( n \to \infty \), and if there exists some estimate \( \delta \) for which \( R(\omega, \delta) \equiv r \), then \( \delta \) is a minimax estimate.

We shall frequently need another theorem in the sequel. We state and prove it as

Theorem 4.3. If \( \delta, r \) be the minimax procedure and the minimax risk respectively, assuming that the observations \( X \) follow any probability distribution \( \omega \in \Omega^* \), and if \( \Omega \supseteq \Omega^* \) be a space of distributions for which the risk associated with \( \delta \) does not exceed \( r \), then \( \delta \) is a minimax procedure and \( r \) the minimax risk for all distributions of \( X \) in \( \Omega \).

Proof. Let \( d \) be any decision procedure. Since \( r \) is the minimax risk for all distributions in \( \Omega^* \),

\[
\sup_{\omega \in \Omega^*} R(\omega, d) \geq r
\]

and hence

\[
(1) \quad \sup_{\omega \in \Omega} R(\omega, d) \geq r
\]

\hfill 1/ \text{Ibid., p. 4-29.}
Also \( S \) is a minimax procedure for all distributions in \( \Omega^* \), i.e.,

\[
\sup_{\omega \in \Omega^*} R(\omega, S) = r
\]

Therefore

(2) \( \sup_{\omega \in \Omega} R(\omega, S) \geq r \)

But we are given that for all distributions \( \omega \in \Omega \),

\[
R(\omega, S) \leq r
\]

so that

(3) \( \sup_{\omega \in \Omega} R(\omega, S) \leq r \)

From (2) and (3) it follows that

(4) \( \sup_{\omega \in \Omega} R(\omega, S) = r \)

Hence from (1) and (4) we see that

\[
\sup_{\omega \in \Omega} R(\omega, d) \geq \sup_{\omega \in \Omega} R(\omega, S) = r
\]

which shows that \( S \) is the minimax procedure and \( r \) the minimax risk for all distributions in \( \Omega \).

V. Bayes and minimax procedures for estimating the proportion defective in a lot of finite size.

Let \( N \) be the size of the lot and \( D \) the number of defective items in it. It is required to estimate the proportion of defective items \( p = D/N \). The statistician is allowed to select \( n \) items out of \( N \) and to observe the number of defective items \( x \) in the sample. The loss function we shall take as

\[
L(p, S) = \frac{(S - p)^2}{p(1-p)}
\]

where \( S \) is the estimate of \( p \).

The problem faced by the statistician is twofold. Firstly, which \( n \) out of the \( N \) items in the lot should he choose and secondly, what function \( S \) should he adopt to estimate \( p \)?
The first part of the problem, as mentioned in Section VII above, is solved by the application of the principle of invariance and the principle of sufficiency. These principles lead to a strategy which tells him to select each possible sample of \( n \) out of \( N \), without regard to order, with probability \( \frac{1}{\binom{N}{n}} \). Thus he chooses the sample of \( n \) at random, and observes those items for quality.

Let \( x_i \) represent the quality of the \( i^{th} \) item inspected and let \( x_i = 1 \) if the item is defective and \( x_i = 0 \) if it is non-defective. The sample space \( Z \) consists of \( 2^n \) points (i.e., sequences of 1's and 0's) and the second part of the problem faced by the statistician is the choice of a function \( \delta \), defined on \( Z \), which specifies for each \( z \in Z \) the real number chosen to estimate \( p \) when that \( z \) is observed.

Let \( x = \sum_{i=1}^{n} x_i \) be the number of defective items found in the sample. It is known that \( x \) is a sufficient statistic for \( p \) and thus we can take \( \delta \) to be a function of \( x \) only.

Next we notice that the loss function \( L \) is not defined for \( p = 0 \) and \( p = 1 \). The only value of \( x \) we can get is zero when \( p = 0 \) and is 1 when \( p = 1 \). Thus the only procedure we can consider, if \( p = 0 \) and \( p = 1 \) are at all possible true proportion defective, is to take \( \delta(0) = 0 \) and \( \delta(1) = 1 \). Similarly, it is clear that the only function which yields finite risk for all \( p \) is one for which \( \delta(0) = 0 \) and \( \delta(1) = 1 \), unless of course the eventualities \( p = 0 \) and \( p = 1 \) are considered impossible.

Now one way of obtaining a minimax estimate is to guess a least favorable a priori distribution for nature. Intuitively, one would surmise it to be the uniform distribution over all possible values of \( p \), viz.,

\[
\lambda(p) = \frac{1}{N+1}, \quad p = 0, \frac{1}{N}, \ldots, \frac{N-1}{N}, 1
\]
An alternative approach for guessing a least favorable distribution would be to go back to the formulation of the problem of sampling from a finite population, given in Section III. Nature has performed a fixed sample size experiment to provide \((x_1, x_2, \ldots, x_N)\) as the quality of the lot. Here \(x_i = 1\) if the item is defective and \(x_i = 0\) if non-defective. In the case of complete ignorance of the statistician's strategy a logical procedure for nature (assuming its interests are opposed to the statistician's) is to assign the value 1 to each \(x_i, i = 1, 2, \ldots, N\) with a probability \(\omega\), the value \(\omega\) having been chosen from a uniform distribution over \((0, 1)\). This would give the distribution of the number of defectives in the lot as
\[
P\left(\sum_{i=1}^{N} x_i = D \mid \omega\right) = \binom{N}{D} \omega^D (1-\omega)^{N-D}, \quad D = 0, 1, \ldots, N.
\]
But the distribution of \(\omega\) is given by
\[
f(\omega) = 1, \quad 0 \leq \omega \leq 1.
\]
Hence we get
\[
P\left(\sum_{i=1}^{N} x_i = D\right) = \int_{0}^{1} \binom{N}{D} \omega^D (1-\omega)^{N-D} f(\omega) d\omega
\]
\[
= \binom{N}{D} \frac{D!(N-D)!}{(N+1)!}, \quad D = 0, 1, \ldots, N
\]
which can be written as
\[
\lambda(p) = \frac{1}{N+1}, \quad p = 0, 1/N, \ldots, N-1/N, 1.
\]

Let us then obtain the Bayes estimate against this a priori distribution of \(p\).

The Bayes estimate \(\hat{\delta}_\lambda\) is that \(\delta\) which minimizes \(E_\delta R(p, \delta)\), the suffix \(\delta\) with \(E\) denoting that \(\delta\) is to be held constant when expectation is being taken. Thus we seek a function \(\delta\) which would minimize
(1) \[ E_{\varepsilon} R(p, S) = E_{\varepsilon} E_{p} L(p, S) \]
\[ = \sum_{p} \left[ \sum_{z} \frac{(S(z)-p)^2}{p(1-p)} \cdot P(z) \right] \lambda(p) \].

Since each term in the summation is non-negative, we can interchange the order of summation and enquire, for each fixed value \( z \), what value \( S(z) \) will minimize the expression

(2) \[ \sum_{p} \frac{(S(z)-p)^2}{p(1-p)} \cdot P(z) \lambda(p) \].

The value \( S_{\lambda}(z) \) which would minimize this is obviously given by

(3) \[ S_{\lambda}(z) = \frac{\sum_{p} \frac{p}{p(1-p)} \cdot P(z) \lambda(p)}{\sum_{p} \frac{1}{p(1-p)} \cdot P(z) \lambda(p)} \].

Now
\[ P(z) = \binom{N-D}{n-z} \binom{D}{z} \], \( z = 0, 1, \ldots, n \)

and \( \lambda(p) = \frac{1}{N+1} \), \( p = 0, 1/N, \ldots, 1 \).

Thus we obtain from (3)

(4) \[ S_{\lambda} = \frac{\sum_{p} \frac{1}{1-p} \binom{N-D}{n-z} \binom{D}{z}}{\sum_{p} \frac{1}{p(1-p)} \binom{N-D}{n-z} \binom{D}{z}} \].

Since \( p = D/N \), we can replace the summation over the whole range of \( p \) in (4) by the summation over the whole range of \( D \). For a given \( x \), the least and the greatest values that \( D \) can possibly have are \( z \) and \( N-n+z \) respectively.

Hence we get from (4)

(5) \[ S_{\lambda} = \frac{\sum_{D = z}^{N-n+z} \frac{N}{N-D} \binom{N-D}{n-z} \binom{D}{z}}{\sum_{D = z}^{N-n+z} \frac{N^2}{D(N-D)} \binom{N-D}{n-z} \binom{D}{z}} \].
\[
(6) \quad \frac{\sum_{D=2}^{N-n+z} \frac{(N-D-1)!D!}{(N-D-n+z)!(D-z)!}}{\sum_{D=z}^{N-n+z} \frac{(N-D-1)!(D-1)!}{(N-D-n+z)!(D-z)!}} = \frac{1}{N} \frac{a}{b}
\]

Now, using the fact that
\[
\sum_{m=0}^{M} \int_{0}^{1} \frac{(\alpha+\beta+1)!}{\alpha!(\beta)!} x^{\alpha}(1-x)^{\beta} \binom{M}{m} x^{m}(1-x)^{M-m} \, dx = 1
\]
we derive the identity
\[
(8) \quad \sum_{m=0}^{M} \frac{(m+\alpha)!(M-m+\beta)!}{m!(\alpha+\beta+1)!} = \frac{(M+\alpha+\beta+1)!\alpha!\beta!}{M!(\alpha+\beta+1)!}
\]
Substituting \(m=D-z\) and \(M=N-n\) in (8) we obtain
\[
(9) \quad \sum_{D=z}^{N-n+z} \frac{(D-z+\alpha)!(N-D+n+z+\beta)!}{(D-z)!(N-D-n+z)!} = \frac{(N-n+\alpha+\beta+1)!\alpha!\beta!}{(N-n)!(\alpha+\beta+1)!}
\]
Substituting \(\alpha = z, \beta = n-z\), in (9) we get
\[
(10) \quad a = \frac{N!z!(n-z-1)!}{(N-n)!n!}
\]
Substituting \(\alpha = z-1, \beta = n-z\) in (9) we get
\[
(11) \quad b = \frac{(N-1)!(z-1)!(n-z-1)!}{(N-n)!(n-1)!}
\]
Substituting now the values of \(a\) and \(b\) from (10) and (11) respectively in (7), we obtain
\[
\xi_{\lambda} = \frac{1}{N} \frac{Nz}{n} = \frac{z}{n}
\]
Hence the Bayes estimate corresponding to the a priori distribution
\[
\lambda(p) = \frac{1}{N+1} \quad , \quad p = 0, 1/N, \ldots, 1
\]
is given by \(\xi_{\lambda} = z/n\) = proportion defective in the sample.

Next, we find the Bayes risk, associated with this estimate, as
\[ R(p, \lambda) = \mathbb{E} L(p, \lambda) \]
\[ = \sum_{z=0}^{n} \frac{(z_n - p)^2}{p(1-p)} \frac{n!}{(n-z)!} \frac{(n-z)!}{n!} \]
\[ = \frac{1}{n^2 p(1-p)} \frac{N-n}{N-1} np(1-p) \]
\[ = \frac{1}{n} \frac{N-n}{N-1}. \]

This is independent of \( p \), and by defining \( L(0,0) = L(1,1) = \frac{N-n}{n(N-1)} \) this risk becomes constant (for all \( p \)), and we conclude from Theorem 4.1 that \( z/n \) is a minimax estimate, while \( \lambda(p) = \frac{1}{N+1} \), \( p = 0, 1/N, \ldots, 1 \), is the least favorable distribution of nature.

VI. Bayes and minimax procedures for estimating the mean of a stratified population.

A. Infinite population.

Stratification is a very common practice in sample surveys. The entire population is divided into a number of strata and sampling is carried out independently in each stratum.

It can be easily shown that if \( c_i \) be the known cost of sampling per unit in the \( i \)th stratum, \( \mu_i \) and \( \sigma_i^2 \) the unknown mean and known variance of the population in the \( i \)th stratum, and \( k \) be the number of strata, then for a given cost \( C = \sum_{i=1}^{k} c_i n_i \) where \( n_i \) is proposed to be the size of the random sample in the \( i \)th stratum, the optimum procedure (which estimates mean with minimum variance) is to choose \( n_i \) proportional to \( \frac{N_i \sigma_i}{c_i} \) and to estimate the total \( \sum_{i=1}^{k} N_i \mu_i \) by \( \sum_{i=1}^{k} N_i \bar{X}_i \) where \( N_i \) is the size of the \( i \)th stratum, and \( \bar{X}_i \) the simple mean of the \( n_i \) observations selected at random from it. The same result holds when for a given variance of the estimate, the object is to minimize the total cost of sampling.
We shall find for this problem Bayes and minimax procedures by first assuming that the $i^\text{th}$ stratum consists of a normally distributed population with unknown mean $\mu_i$ and known variance $\sigma_i^2$, $i = 1, 2, \ldots, k$; and we shall estimate a linear function of $\mu_i$'s, viz.,

$$u = \sum_{i=1}^{k} \frac{N_i}{n_i} \mu_i$$

where $N_i$ are some given constants.

As pointed out in the introduction, it is desirable that the losses resulting from the errors in the estimates and from the cost of sampling be considered together, and so we shall take loss function as

$$L(u, \delta) = (\delta - u)^2 + \sum_{i=1}^{k} c_i n_i$$

where $n_i$ is the size of the sample chosen from the $i^\text{th}$ stratum, $c_i$ the sampling cost per unit in that stratum, and $\delta$ is a function of the sample point

$$X = (X_{11}, \ldots, X_{kn_1}; X_{21}, \ldots, X_{kn_2}; \ldots, X_{k1}, \ldots, X_{kn_k})$$

where the coordinate $X_{ij}$ stands for the observation $j$ from the $i^\text{th}$ stratum, $(i = 1, \ldots, k, j = 1, \ldots, n_i)$.

It may be noted that a slightly more realistic loss function would be

$$L(u, \delta) = a(\delta - u)^2 + \sum c_i n_i$$

where $a$ is some constant depending upon the desired relative accuracy of results and the cost of experimentation in a given situation. But it is easily seen that any procedure corresponding to this loss function can be obtained from the corresponding procedure in the previous case simply by substituting $c_i/a$ for $c_i$. The risk associated with it will be a times the corresponding risk in the previous case.
The method of finding the minimax estimate used in the previous section will not work here, since there is no least favorable distribution of nature. What we would want is a uniform distribution over the real line but this is not a distribution. We shall make use of Theorem 4.2 and find Bayes solutions corresponding to the a priori distributions $\Lambda = N(0, \theta^2)$, and by letting $\theta \to \infty$. Then if the Bayes risks $\gamma \theta$ corresponding to the Bayes estimates $\delta \theta$ would tend to a constant limiting value, say $r$, then any procedure which has this constant $r$ as its risk will be a minimax procedure.

Since the sample mean of each stratum is a sufficient estimate for the corresponding stratum mean, it follows that we can take $\delta$ to be a function of $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k$ where $\bar{x}_i$ is the sample mean from the $i^{th}$ stratum.

We can regard $n_1, n_2, \ldots, n_k$ as fixed for the purpose of finding the estimates, and later choose them so as to minimize the total risk which would be obtained as a function of $n_1$'s. Let now $X$ stand for $(\bar{x}_1, \ldots, \bar{x}_k)$ and $\mu$ for $(\mu_1, \ldots, \mu_k)$.

$$R(u, \delta) = E_\mu[(\delta-u)^2 + \sum c_i n_i]$$

Bayes estimate $\delta \theta$ is that function $\delta$ which minimizes the average risk computed with respect to the a priori probability density of $\mu_1$. Thus we require a function $\delta$ which would minimize

$$E_\delta E_\mu[(\delta-u)^2 + \sum c_i n_i] = E_\mu E_\delta[(\delta-u)^2 + \sum c_i n_i]$$

the change in the order of taking expectations being valid as the integrand is non-negative. The problem then reduces to the finding of the value $\delta(x)$ for each value $x$ of $X$, so that

$$\int [(\delta(x)-u)^2 + \sum c_i n_i] p(x) d \lambda(\mu)$$

is minimized, $d \lambda(\mu)$ standing for $\prod_{i=1}^k d \lambda(\mu_i)$. This value $\delta(x)$ is easily seen to be
\[ S(x) = E(u|X=x) \]

and the corresponding Bayes risk is obtained as
\[ E_u[\sigma^2 u|X + \sum c_i n_i] \]

If the term in the square bracket is independent of \( X \), then of course it is the Bayes risk.

Since \( \mu_i \)'s have a prior distribution
\[ \lambda_0(\mu_i) = N(0, \theta^2) \]

and they are independent, we get
\[ p(\mu_i) = \text{const. e} \]

\[ -\frac{1}{26^2} \sum_{i=1}^{k} \mu_i^2 \]

Since we have assumed \( X_{ij} \) to be normally distributed for each \( i \), \( \bar{x}_i \) is also normally distributed with mean \( \mu_i \) and variance \( \sigma_i^2/n_i \), so that
\[ p(x|\mu) = \text{const. e} \]

\[ -\frac{1}{2} \sum_{i=1}^{k} \frac{x_i - \mu_i}{\sigma_i}^2 \]

We therefore obtain
\[ p(\mu|x) = c(x), e \]

\[ -\frac{1}{2} \sum_{i=1}^{k} \left[ \frac{n_i}{\sigma_i^2} \frac{x_i - \mu_i}{\sigma_i^2} + \frac{\mu_i^2}{\theta^2} \right] \]

\[ = c(x), e \]

where
\[ y_i = \frac{n_i \bar{x}_i}{\sigma_i^2} \left( \frac{n_i}{\sigma_i^2} + \frac{1}{\theta^2} \right) \]

Since the joint distribution of \( \mu_1, \mu_2, \ldots, \mu_k \) for given \( x \) factors into \( k \) normal distributions, we find that the distributions of \( \mu_i \) for given \( x \) are independent normal with mean \( y_i \) and variance \( \frac{n_i}{\sigma_i^2} + \frac{1}{\theta^2} \). Thus, for given \( x \),
the distribution of \( u = \sum N_i \mu_i \) is normal with mean \( \sum N_i \theta_i \) and variance \( \sum \frac{N_i^2}{n_i + \frac{1}{\theta_i^2}} \). We thus conclude that the Bayes estimate \( \hat{\theta} = \sum N_i \theta_i \), and since \( \sigma^2 u / X \) is independent of \( X \), we get the Bayes risk
\[
r_\theta = \sum \frac{N_i^2 \sigma_i^2 \theta_i^2}{n_i \theta_i^2 + \sigma_i^2} + \sum c_i n_i.
\]
Now let \( \theta \to \infty \), and we see that
\[
r_\theta \to r = \sum \left( \frac{N_i^2 \sigma_i^2}{n_i} + c_i n_i \right).
\]
Now if we could find some estimate \( \Delta \) which has a constant risk \( r \), then that \( \Delta \) would be a minimax estimate by virtue of Theorem 4.2. Let us try the limiting estimate \( \lim_{\theta \to \infty} \hat{\theta} = \sum N_i \bar{X}_i \).

\[
R(u, \sum N_i \bar{X}_i) = E(\sum N_i \bar{X}_i - u)^2 + \sum c_i n_i.
\]

Since \( \bar{X}_i \) is normal with mean \( \mu_i \), variance \( \frac{\sigma_i^2}{n_i} \); \( \sum N_i \bar{X}_i \) is also normal with mean \( \sum N_i \mu_i = u \) and variance \( \sum N_i^2 \sigma_i^2 / n_i \). Hence the risk corresponding to the estimate \( \sum N_i \bar{X}_i \) is given by
\[
R = \sum \left( \frac{N_i^2 \sigma_i^2}{n_i} + c_i n_i \right) = r.
\]
Thus we conclude that for given \( n_i \)'s, \( \sum N_i \bar{X}_i \) is a minimax estimate, with risk \( r \). Now we would want to choose \( n_i \) such that for all positive integral values of \( n_i \), \( r \) is minimum. Since each term in \( r \) is non-negative, we minimize
\[
\frac{N_i^2 \sigma_i^2}{n_i} + c_i n_i
\]
for proper choice of \( n_i \). If we denote this expression by \( r_i(n_i) \) we get
\[
r_i(n_i + 1) = \frac{N_i^2 \sigma_i^2}{n_i + 1} + c_i(n_i + 1)
\]
Hence
\[ r_i(n_{i+1}) - r_i(n_i) = \frac{N_i^2 \sigma_i^2}{n_i(n_{i+1})}. \]

To minimize \( r_i(n_i) \), we choose that value for \( n_i \) for which this difference changes first from negative to positive, in other words for which \((n_i + \frac{1}{2})^2\) first exceeds \( N_i^2 \sigma_i^2 / (c_i) + \frac{1}{4} \). This gives
\[ n_i = \text{integer nearest to} \sqrt{\frac{N_i^2 \sigma_i^2}{c_i} + \frac{1}{4}}. \]

When \( \sqrt{\frac{N_i^2 \sigma_i^2}{c_i} + \frac{1}{4}} \) lies exactly between two integers, say \( m \) and \( m+1 \), the risk is equal and minimum for both \( n_i = m \) and \( n_i = m+1 \) and it is immaterial which of the two nearest integers we choose for \( n_i \).

**Removal of the assumption of normality.** We obtained the minimax estimate above for \( u = \sum N_i \mu_i \) assuming that the observations are normally distributed in each stratum. We shall now use Theorem 4.3 to remove this restriction.

Let us only assume that whatever the distribution in each stratum, they are independent, and the distribution in the \( i \)th stratum has a known variance \( \sigma_i^2 \), \((i = 1, 2, \ldots, k)\). Let us calculate the risk \( R \) corresponding to the minimax estimate \( \hat{\delta} \) obtained for normal distribution of the observations in each stratum, under these circumstances.

\[ R = E(\sum_{i=1}^{k} N_i \bar{X}_i - u)^2 + \sum c_i n_i \]

\[ = E[\sum N_i (\bar{X}_i - \mu_i)^2] + \sum c_i n_i \]

\[ = \sum_{i=1}^{k} N_i E(\bar{X}_i - \mu_i)^2 + \sum_{i \neq h} N_i N_h E(\bar{X}_i - \mu_i)(\bar{X}_h - \mu_h) + \sum c_i n_i. \]

But the distributions in the strata are independent, and
\[ E(\bar{X}_i) = \mu_i, \quad E(\bar{X}_i - \mu_i)^2 = \sigma_i^2 / n_i. \]
Thus the middle term vanishes, and we get

\[ R = \sum \left( \frac{N_i^2 \sigma_i^2}{n_i} + c_i n_i \right) = \frac{r}{n_i} \]

Applying Theorem 4.3 now, we conclude that the minimax procedure found under the assumption of normality of observations, viz., \( \delta = \sum N_i \bar{x}_i \) is still minimax when nothing is assumed about the distribution of the observations excepting the independence from stratum to stratum, and their variance.

B. **Finite population.**

Suppose now that \( X_{ij} \), where \( i = 1, \ldots, k; j = 1, \ldots, N_i \), denotes some characteristic of the \( j^{th} \) unit in the \( i^{th} \) stratum. Suppose further that \( N_i \) is known and that we are required to estimate the total

\[ u = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X_{ij} = \sum_{i=1}^{k} N_i u_i \]

for the entire population and we want to adopt the principle of minimax estimation, the loss function being the same as before. In terms of the formulation of the problem of sampling from a finite population given in Section III, we look at the finite populations in the different strata as if they were the result of performance of a fixed sample size experiment (\( N_i \) in the \( i^{th} \) stratum) by nature or some conscious human being, the experiments being carried out independently for the various strata. The statistician does not know the probability density in any case beyond the fact that the distribution in the \( i^{th} \) stratum has variance \( \sigma_i^2 \; i = 1, \ldots, k \).

He is allowed to choose a fixed number of units from the entire \( N = \sum N_i \) units in any manner he likes and observe them to find the corresponding values for \( X_{ij} \).
It has been pointed out in Section IV that for any finite population, which can be regarded as having been obtained by a single experiment by nature, the best strategy for the statistician is to adopt simple random sampling. In this case he is given the variance in the different strata and he can very well proceed on the assumption that each experiment was carried out independently, possibly with a different distribution each time. In any case, it is obvious that the best strategy for him is to choose some fixed number, say \( n_i \), from the \( i \)th stratum by simple random sampling and to do so for each stratum. The number \( n_i \) would, of course, be obtained by minimizing the total minimax risk obtainable for some set of fixed \( n_i \)'s.

We shall first assume that the distribution which gave rise to the point \( X = (X_{i1}, X_{i2}, \ldots, X_{iN_i}) \) was normal, with unknown mean \( \mu_i \) and known variance \( \sigma_i^2 \). Supposing that somehow the statistician has decided to take a sample of size \( n_i \). Since the entire population out of which he is selecting his sample consists of \( N_i \) elements \( X_{ij}, j = 1, \ldots, N_i \), and he is asked to estimate the total \( \mu = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X_{ij} \), his sampling amounts to sampling from a normal population with \( \sum_{j=1}^{N_i} X_{ij} \) fixed, and equal to \( N_i \mu_i \).

As in the case of infinite population, the sample mean from each stratum is a sufficient statistic for the corresponding stratum mean, we can take our \( S \) to be a function of \( \overline{X}_1, \overline{X}_2, \ldots, \overline{X}_k \), where \( \overline{X}_i \) is the sample mean from \( i \)th stratum.

The joint distribution of \( X_{i1}, X_{i2}, \ldots, X_{in_i} \) where \( \sum_{j=1}^{n_i} X_{ij} \) is fixed and \( = N_i \mu_i \), is easily seen to be \( n_i \)-variate normal, with variance-covariance matrix as
\[ \sigma_i^2 \left( \frac{N_i-1}{N_i}, \ldots, \frac{1}{N_i} \right) \]

The joint distribution of \( \bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \) will be seen to be normal with mean \( \mu_i = \mu \) and variance

\[ \frac{1}{2} \left[ \frac{n_i(N_i-1)}{n_i N_i} - n_i(n_i-1) \right] \frac{1}{N_i} \sigma_i^2 = \frac{N_i-n_i}{n_i N_i} \sigma_i^2 = v_i, \text{ say.} \]

Therefore,

\[ p(x|\mu) = \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{k} \frac{v_i}{x_i-u_i}^2} \]

\( \mu \) standing jointly for \( u_1, u_2, \ldots, u_k \) and \( x \) for \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \). Now

\[ -\frac{1}{2} \sum_{i=1}^{k} \frac{v_i}{u_i^2} \]

\[ p(\mu) = \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{k} \frac{(\bar{x}_i-u_i)^2 + u_i^2}{v_i}} \]

Therefore

\[ p(\mu|x) = c(x) \cdot e^{-\frac{1}{2} \sum_{i=1}^{k} \frac{(\frac{1}{v_i} + \frac{1}{\sigma_i^2})(u_i-y_i)^2}{v_i}} \]

where

\[ y_i = \frac{x_i}{v_i} \left( \frac{1}{v_i} + \frac{1}{\sigma_i^2} \right) \]

Since the joint distribution of \( u_1, \ldots, u_k \) for given \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \) factors, we find that the conditional distributions of \( u_i \) for given \( \bar{x} \) are independent and normal with mean \( y_i \) and variance \( 1/(v_i^2 + \sigma_i^2) \). Thus for given \( \bar{x} \), the distribution of \( u = \sum N_i u_i \) is normal with mean \( \sum N_i y_i \) and variance \( \sum \frac{N_i^2}{v_i + \sigma_i^2} \). We thus conclude that the Bayes estimate \( \delta_\theta = \sum N_i y_i \), and the
Bayes risk \( r_\theta = \sum \frac{N_i^2 v_i \theta^2}{\theta^2 + v_i} + \sum c_i n_i \). Now let \( \theta \to \infty \), then

\[ r_\theta \to r = \sum (N_i^2 v_i + c_i n_i) = \sum \left( N_i^2 \frac{n_i - n_i}{n_i N_i} \sigma_i^2 + c_i n_i \right) \]

Now an estimate \( \delta \) having the constant risk \( r \) will be, by Theorem 4.2, a minimax estimate. Let us try the limiting Bayes estimate \( \lim \delta_\theta = \sum \frac{N_i \bar{x}_i}{N_i} \to \infty \)

Since \( \bar{x}_i \) is seen already to be normal with mean \( u_i \) and variance \( v_i \), it follows that \( \sum N_i \bar{x}_i \) is also normal with mean \( \sum N_i u_i = u \) and variance \( \sum N_i^2 v_i \). Hence the risk corresponding to the estimate \( \sum N_i \bar{x}_i \) is given by

\[ R = \sum (N_i^2 v_i + c_i n_i) \]

\[ = \sum (N_i^2 \frac{n_i - n_i}{n_i N_i} \sigma_i^2 + c_i n_i) = r \]

Thus we conclude that for finite populations too, for given \( r_i, \sum \frac{N_i \bar{x}_i}{N_i} \) is a minimax estimate with risk \( r \). As before we now choose positive integral values for \( n_i \leq N_i \), such that \( r \) is minimized. Each term in \( r \) being non-negative, and the \( i \)th term \( r_i(n_i) \) being independent of \( n_h, h \neq i \), we get \( n_i \) by minimizing \( r_i(n_i) \) only.

\[ r_i(n_i) = N_i^2 \frac{n_i - n_i}{n_i N_i} \sigma_i^2 + c_i n_i \]

\[ = N_i^2 \left( \frac{1}{n_i} - \frac{1}{N_i} \right) \sigma_i^2 + c_i n_i \]

\[ r_i(n_i + 1) = N_i^2 \left( \frac{1}{n_i + 1} - \frac{1}{N_i} \right) \sigma_i^2 + c_i(n_i + 1) \]

Therefore,

\[ r_i(n_i + 1) - r_i(n_i) = c_i - \frac{N_i^2 \sigma_i^2}{n_i(n_i + 1)} \]
Exactly as before, we get

\[ n_i = \text{integer nearest to } \sqrt{\frac{N_i^2 \sigma_i^2}{c_i} + \frac{1}{4}} \]

but this time with the restriction that \( n_i \leq N_i \). In case the value for \( n_i \) is found to be bigger than \( N_i \), it has to be taken equal to \( N_i \).

Before trying to remove the assumption of normality with the help of Theorem 4.3, let us first consider how we shall define the variance for a finite population. We see that the conditional distribution of \( X_{ij} \), given \( \sum_{j=1}^{N_i} X_{ij} = N_i u_i \) is normal with mean \( u_i \) and variance \( \frac{N_i - 1}{N_i} \sigma_i^2 \) when the \( X_{ij} \)'s are distributed normally for any \( i \). It can, however, be seen easily that whatever be the conceptual distribution responsible for the existence of the finite population of size \( M \), say \( Y_1, Y_2, \ldots, Y_M \), but having mean \( u \) and variance \( \sigma^2 \), the conditional distribution of a random sample of size \( m \) (say of \( Y_1, \ldots, Y_m \)), given \( Y_1 + \ldots + Y_M = Mu \), with \( m \leq M \), will be some \( m \)-variate distribution, with the following properties, where \( y \) stands for the column vector with \( m \) components \( (Y_1, \ldots, Y_m) \) and \( u \) stands for the column vector with \( m \) components \( (u, u, \ldots, u) \):

\[
E(y | Y_1 + \ldots + Y_M = Mu) = u
\]

\[
E(y - u)(y - u)' = ||a_{ij}||
\]

where

\[
a_{ij} = \frac{M-1}{M} \sigma^2 \quad \text{if } i = j
\]

\[
= -\frac{1}{M} \sigma^2 \quad \text{if } i \neq j
\]

\[ i,j = 1,2,\ldots,m \]

But we can also calculate directly from the finite population the following well known results on the basis of sampling from a finite population without replacement:
\[ E(Y_i) = \frac{1}{M} \sum_{i=1}^{M} Y_i \]
\[ E[Y_i - E(Y_i)]^2 = \frac{1}{M} \sum_{i=1}^{M} [Y_i - E(Y_i)]^2 \]
\[ E[Y_i - E(Y_i)][Y_j - E(Y_j)] = -\frac{1}{M(M-1)} \sum_{i=1}^{M} [Y_i - E(Y_i)]^2 \]

Comparing the two sets of results it follows that we should have

\[ u = \frac{1}{M} \sum_{i=1}^{M} Y_i \]

and

\[ \sigma^2 = \frac{1}{M-1} \sum_{i=1}^{M} [Y_i - \frac{1}{M} \sum_{i=1}^{M} Y_i]^2 \]

We shall define the variance of any finite population \( Y_1, \ldots, Y_M \) by simply \( \sigma^2 = \frac{1}{M-1} \sum_{i=1}^{M} (Y_i - u)^2 \) where \( u = \frac{1}{M} \sum_{i=1}^{M} Y_i \), and this variance would, in fact, by the variance of the conceptual underlying distribution used by nature (or some conscious being) to produce \( Y_1, Y_2, \ldots, Y_M \).

**Removal of the normality assumption for the underlying distribution.**

Now suppose we do not assume anything about the distribution used by nature to produce our finite population, beyond the fact that they are independent with unknown means \( u_i = \frac{1}{N_i} \sum_{j=1}^{N_i} X_{ij} \) and known variances

\[ \sigma_i^2 = \frac{1}{N_i-1} \sum_{j=1}^{N_i} (X_{ij} - u_i)^2 , \quad i=1, \ldots, k \]

This is in accordance with the assumption for stratified sampling, that the variance in each stratum is known. Let us now calculate the risk \( R \) for the estimate \( \delta = \sum_{i=1}^{k} N_i \bar{X}_i \) without assuming normality of the underlying distribution.

\[ R = E_u[(\delta - u)^2] = \sum_{i=1}^{k} \sum_{i=1}^{n_i} c_{i,n_i} \]
\[ = E\left( \sum_{i=1}^{k} N_i \bar{X}_i - \sum_{i=1}^{k} N_i u_i \right)^2 + \sum_{i=1}^{k} \sum_{n_i} c_{i,n_i} \]
\[ = E\left[ \sum_{i=1}^{k} N_i^2 (\bar{X}_i - u_i)^2 + \sum_{i \neq h} N_i N_h (\bar{X}_i - u_i)(\bar{X}_h - u_h) + \sum_{i=1}^{k} \sum_{n_i} c_{i,n_i} \right]. \]
But sampling is done independently in the various strata and \( E(X_i - u_i) = 0 \), so the middle term vanishes and we are left with

\[
R = \sum_i [N_i^2 E(X_i - u_i)^2 + c_i n_i]
\]

But

\[
E(X_i - u_i)^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (X_{ij} - u_i)^2
\]

\[
= \frac{1}{n_i} \left[ \sum_{j=1}^{n_i} E(X_{ij} - u_i)^2 + \sum_{j \neq h} E(X_{ij} - u_i)(X_{ih} - u_i) \right]
\]

\[
= \frac{1}{n_i} \left[ \frac{n_i}{N_i} (N_i - 1) \sigma_i^2 - \frac{n_i(n_i - 1)}{N_i(N_i - 1)} (N_i - 1) \sigma_i^2 \right]
\]

\[
= \frac{N_i - n_i}{n_i N_i} \sigma_i^2
\]

Hence

\[
R = \sum_i \left( N_i^2 \frac{N_i - n_i}{n_i N_i} \sigma_i^2 + c_i n_i \right) = r
\]

Thus we see that the risk is the same when we do not assume any distribution for \( X_i \)'s. Applying Theorem 4.3, we conclude that when \( N_i \) and \( \sigma_i^2 = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (X_{ij} - u_i)^2 \) are given for every stratum, and \( \sum_{i=1}^{k} \sum_{j=1}^{N_i} X_{ij} \)

\( \sum_{i=1}^{k} N_i u_i \) is to be estimated, the usual estimate \( \sum_{i=1}^{k} N_i \bar{X}_i \) is a minimax estimate and the optimum size of the sample (for minimax sampling) is given by

\[
n_i = \text{integer nearest to } \sqrt{\frac{N_i^2 \sigma_i^2}{c_i} + \frac{1}{4}}
\]

and not exceeding \( N_i \), the cost of experimentation being \( c_i \) per sampling unit in the \( i \)th stratum. The minimax risk is given by

\[
r = \sum_{i=1}^{k} \left[ N_i^2 \left( \frac{1}{N_i} - \frac{1}{N_i} \right) \sigma_i^2 + c_i n_i \right]
\]
As mentioned earlier about the relative importance of the error in the estimate and sampling cost, we can say that our results indicate the following:

"If the statistician were to be penalized at the rate of 'a' dollars per unit of the squared error in the estimate (i.e., the square of the difference in the estimate and true unknown value) and it would cost him 'c_i' dollars per sampling unit in the i^{th} stratum for experimentation, then a minimax procedure for him would be to sample at random from each stratum, a number n_i of the sampling units, determined by

\[ n_i = \text{integer nearest to } \sqrt{\frac{an_i^2 \sigma_i^2}{c_i} + \frac{1}{4}} \]

and not exceeding N_i, and use \( \sum N_i \bar{x}_i \) as his estimate for the total \( \sum N_i u_i \). His expected loss would not exceed

\[ \sum [aN_i^2 \left( \frac{1}{n_i} - \frac{1}{N_i} \right) \sigma_i^2 + c_i n_i] \]

dollars in any case."

Note that the loss function for this case is

\[ L(u, \delta) = a(\delta - u)^2 + \sum c_i n_i \]

VII. Bayes and minimax procedures for estimating mean of a population with two-stage (cluster) sampling.

Two stage sampling in sample surveys is the procedure under which, instead of taking samples from each stratum as discussed in stratified sampling, a sample at random is first chosen from all the strata, considering a stratum as the unit of sampling. Then a simple random sample is chosen from each of the strata obtained in the sample. Thus sampling is carried out in two stages, the strata being the first-stage sampling units and the sampling units composing each stratum being the second-stage sampling units. In case the sampling ends with the sampling of strata and each stratum obtained in the sample is
enumerated completely, the process is termed "cluster sampling", the name
cluster being used for stratum.

Thus if we consider the whole finite population of size MN, say, divided
into N first-stage sampling units, each first stage unit consisting of M
second-stage sampling units, and a sample of n first-stage units at random
is taken, and m second-stage units sampled at random from each of the n
first-stage units in the sample, then

if \( N = n, M = m \) : it is called "complete enumeration".
if \( N = n, M < n \) : " " " stratified sampling"
if \( n < N, M = m \) : " " " cluster sampling"
and if \( n < N, M < m \) : " " " two-stage sampling" .

We shall in this section consider the problem of two-stage sampling and
later obtain the corresponding results for cluster sampling. The problem will
be formulated and solved first for infinite populations and then for finite
populations. We shall, for simplicity, call the first-stage sampling units
by the name of 'clusters'.

A. Infinite populations.

1. Formulation of the problem.

Consider a situation where a statistician is required to estimate the
mean \( \bar{u} \) of some distribution with variance \( \sigma_b^{-2} \). He is allowed to choose a
random sample of some predetermined size, say \( n \), but not allowed to observe
the values \( u_1, u_2, \ldots, u_n \), obtained. He is, however, allowed to observe a
predetermined number of observations, say \( m \) from each of the conditional
distributions of another random variable \( X \) corresponding to the values
\( u_1, \ldots, u_n \), obtained and unobserved by him, the conditional distributions
being such that

\[
E(X|u_1) = u_1, \quad E[(X-u_1)^2|u_1] = \sigma_\omega^2.
\]
He is told that $\sigma_\omega^2$ is the same for each such distribution and is further given the values of both $\sigma_b^2$ and $\sigma_\omega^2$. The loss function being

$$L(\vec{u}, f) = (f-\vec{u})^2 + c_1 nm + c_2 n$$

where $f$ is the function used by the statistician to estimate $\vec{u}$. The problem is, what function $f$ should be adopted to estimate $\vec{u}$, and further what values of $n$ and $m$ should be decided for the sample sizes. In the above, $c_1$ is the sampling cost per unit of observations on $X$ and $c_2$ is the sampling cost per unit for choosing the sample $(u_1, \ldots, u_n)$.

2. Bayes estimate.

As before, we shall first hold $m$ and $n$ as fixed and determine what estimate he should choose for Bayes and minimax estimates. Later we shall choose $m$ and $n$ such that the minimax risk for given $m$ and $n$ is minimized over all possible choices of $m$ and $n$.

It is known that each $\bar{x}_i$ is sufficient for the corresponding $u_i$, and (since the sample $(u_1, \ldots, u_n)$ is sufficient for estimating $\vec{u}$) we can take $f$ to be a function of the means $\bar{x}_1, \ldots, \bar{x}_n$.

Again, there does not exist a least favorable distribution, and we shall find a sequence of Bayes solutions when the a priori probability density of $\vec{u}$ is normal with mean zero and variance $\sigma^2$. Let $X$ stand for $(\bar{x}_1, \ldots, \bar{x}_n)$ and $u$ for $(u_1, \ldots, u_n)$. As in the previous cases, the Bayes estimate will be $E(\vec{u}|X)$ and the Bayes risk will be $E(\sigma^2_{\vec{u}|X}) + c_1 nm + c_2 n = \sigma^2_{\vec{u}|X} + c_1 nm + c_2 n$ if $\sigma^2_{\vec{u}|X}$ is independent of $X$.

We shall first assume that $X$'s are normally distributed, as well as $u$'s. Later we shall see how to do away with these restrictions. We have
\[
p(x|u, \bar{u}) = \text{const.} \prod_{i=1}^{n} \frac{1}{\sigma_\omega^2} e^{-\frac{1}{2} \frac{m(x_i-u_i)^2}{\sigma_\omega^2}}
\]

\[
p(u|\bar{u}) = \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{n} \frac{(u_i-\bar{u})^2}{\sigma_b^2}}
\]

therefore

\[
p(x,u|\bar{u}) = \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{n} \left[ -\frac{m}{\sigma_\omega^2} (\bar{x}_i-u_i)^2 + \frac{1}{\sigma_b^2} (u_i-\bar{u})^2 \right]}
\]

\[
= \text{const.} e
\]

Integrating out \( u_1, u_2, \ldots, u_n \), we get

\[
p(x|\bar{u}) = \text{const.} \prod_{i=1}^{n} \int \exp \left\{ -\frac{1}{2} \left[ \left( \frac{m}{\sigma_\omega^2} + \frac{1}{\sigma_b^2} \right) (u_i - \bar{u})^2 + \frac{m\bar{x}_i}{\sigma_\omega^2} \bar{u} + \frac{u_i}{\sigma_b^2} \bar{u} \right] \right\} du_i
\]

\[
= \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{n} \frac{m(x_i-u_i)^2}{m \sigma_b^2 + \sigma_\omega^2}}
\]

A priori distribution for \( \bar{u} \) is taken as

\[
p(\bar{u}) = \text{const.} e^{-\frac{1}{2} \frac{(-\bar{u})^2}{\sigma_b^2}}
\]

Hence, we get
\[ p(u|X) = c(X) e^{-\frac{1}{2} \left[ \sum \frac{m(x_i - u)^2}{\sigma_b^2 + \sigma_\omega^2} + \frac{u^2}{\sigma^2} \right]} \]

\[ = c(X) e^{-\frac{1}{2} \left[ u^2 \left( \frac{mn}{m \sigma_b^2 + \sigma_\omega^2} + \frac{1}{\sigma^2} \right) - 2u \left( \frac{\sum m \bar{x}_i}{m \sigma_b^2 + \sigma_\omega^2} \right) + \sum \frac{m \bar{x}_i^2}{m \sigma_b^2 + \sigma_\omega^2} \right]} \]

\[ = c(X) e^{-\frac{1}{2r} (u - f_\theta)^2} \]

where

\[ f_\theta = \frac{\sum m \bar{x}_i}{mn + \frac{m \sigma_b^2 + \sigma_\omega^2}{\sigma^2}} \]

and

\[ r_\theta = \frac{1}{mn + \frac{1}{m \sigma_b^2 + \sigma_\omega^2} + \frac{1}{\sigma^2}} \]

Since \( f_\theta = E(u|X) \), we see that it is a Bayes estimate, and since \( r_\theta \) is independent of \( X \), we see that

\[ \text{Bayes risk} = \frac{1}{mn + \frac{1}{m \sigma_b^2 + \sigma_\omega^2} + \frac{1}{\sigma^2}} + c_1 nm + c_2 n \]


Now we shall apply Theorem 4.2 for obtaining a minimax estimate if one exists. Limit of the Bayes risk as \( \theta \to \infty \), is

\[ r = \frac{m \sigma_b^2 + \sigma_\omega^2}{mn + c_1 nm + c_2 n} \]

This is constant, and if we can find an estimate with its risk = \( r \), then that would be minimax.
Let us try
\[ \lim_{\theta \to \infty} f_\theta = f = \frac{1}{mn} \sum_{i=1}^{n} \bar{m}_{i} = \bar{x}_{..}, \quad \text{say}. \]

In finding the risk corresponding to the estimate \( f = \bar{x}_{..} \), we shall not assume any distribution of either \( X \)'s or \( u \)'s. If we then could get the risk constant and equal to \( r \), we shall by virtue of Theorem 4.3 conclude that \( \bar{x}_{..} \) is minimax, whatever be the distributions of \( X \) and \( u \), except that their variances be \( \sigma^2_{\omega} \) for each given \( u \) and \( \sigma^2_{b} \) respectively.

Let \( R \) be the risk for \( f \), then
\[ R = E[(\bar{x}_{..} - \bar{u})^2 + c_{1}nm + c_{2}n]. \]

But
\[ E(\bar{x}_{..} - \bar{u})^2 = E(\frac{1}{n} \sum_{i=1}^{n} x_{i} - \bar{u})^2 \]
\[ = \frac{1}{n} E[\sum_{i=1}^{n} (x_{i} - \bar{u})^2 + \sum_{i \neq j} (x_{i} - \bar{u})(x_{j} - \bar{u})]. \]

But
\[ E(\bar{x}_{i} - \bar{u})(\bar{x}_{j} - \bar{u}) = E(\bar{x}_{i} - u_{i} + u_{i} - \bar{u})(\bar{x}_{j} - u_{j} + u_{j} - \bar{u}) \]
and since the sampling for \( X \)'s and \( u \)'s have been independent, we see that this term vanishes, and since
\[ \sum (\bar{x}_{i} - \bar{u})^2 = \sum (x_{i} - u_{i} + u_{i} - \bar{u})^2 = \sum (\bar{x}_{i} - u_{i})^2 + \sum (u_{i} - \bar{u})^2 \]
\[ + 2 \sum (\bar{x}_{i} - u_{i})(u_{i} - \bar{u}) \]

we obtain
\[ E \sum (\bar{x}_{i} - \bar{u})^2 = \frac{\sigma^2_{\omega}}{m} + n \sigma^2_{b} \]

thus giving
\[
R = \frac{1}{n} \left[ n \frac{\sigma_w^2}{m} + n \sigma_b^2 \right] + c_1 n m + c_2 n
\]
\[
= \frac{\sigma_b^2}{n} + \frac{\sigma_w^2}{m n} + c_1 n m + c_2 n
\]
\[
= r
\]

Hence the estimate \( \tilde{X} \) is a minimax estimate, whatever the distribution be of \( X \) and of \( u = E(X|u) \).

The optimum values of \( m \) and \( n \) (for minimax sampling) can be obtained by choosing \( m \) and \( n \) so as to minimize \( r \).

An approximate solution is obtained by treating \( m \) and \( n \) as if they were continuous variables and obtaining minimum value of \( r \) by the method of calculus. Then we could take \( m \) and \( n \) as the integers nearest to those values, for an approximate solution. Differentiating partially with respect to \( m \) and \( n \) and equating to zero, we obtain the following values for \( m \) and \( n \):

\[
n = \frac{\sigma_b}{\sqrt{c_2}}
\]
\[
m = \sqrt{\frac{c_2}{c_1}} \frac{\sigma_w}{\sigma_b}
\]

It would be seen that the optimum values of \( m \) and \( n \) are those for which the two components of risk, one due to the squared error in the estimate and the other due to sampling costs assume equal values or as nearly equal as the integral values of \( m \) and \( n \) will allow.

4. **Minimax risk in terms of intraclass correlation.**

We shall now express the risk function in terms of intraclass correlation.

First of all, notice that if \( u \) is normally distributed with mean \( \bar{u} \) and variance \( \sigma_u^2 \), and for a given \( u \), if \( X \) is normally distributed with mean \( u \) and variance \( \sigma_w^2 \), then the distribution of \( X \) is normal with mean \( \bar{u} \) and variance \( \sigma_b^2 + \sigma_w^2 \). This is easily seen thus:
- 34 -

\[ f(x|u) = \frac{1}{\sqrt{2\pi} \sigma_\omega} e^{-\frac{1}{2\sigma_\omega^2} (x-u)^2} \]

\[ g(u) = \frac{1}{\sqrt{2\pi} \sigma_b} e^{-\frac{1}{2\sigma_b^2} (u-u)^2} \]

Thus

\[ h(x,u) = \frac{1}{2\pi \sigma_\omega \sigma_b} e^{-\frac{1}{2} \left[ \frac{(x-u)^2}{\sigma_\omega^2} + \frac{(u-u)^2}{\sigma_b^2} \right]} \]

and

\[ p(x) = \int_{-\infty}^{\infty} h(x,u) du \]

\[ = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2\sigma^2} (x-u)^2} \]

where \( \sigma^2 = \sigma_b^2 + \sigma_\omega^2 \).

In fact, we may consider the observation \( x_{i\alpha} \) where \( i \) denotes that \( \text{E}(x_{i\alpha} | i) = u_i \) as made up of two independently distributed chance variables,

\[ x_{i\alpha} = \eta_i + \epsilon_{i\alpha} \]

where

\[ \eta_i : N(\mu, \sigma_b^2) \]

\[ \epsilon_{i\alpha} : N(0, \sigma_\omega^2) \]

Now, intraclass correlation is defined as the simple correlation coefficient between two observations in the same class, i.e., if we consider

\[ x_{i\alpha} = \eta_i + \epsilon_{i\alpha} \]

\[ x_{i\beta} = \eta_i + \epsilon_{i\beta} \]

then the intraclass correlation coefficient \( \rho \) is defined as the simple correlation coefficient between \( x_{i\alpha} \) and \( x_{i\beta} \). This is easily seen to work out to
\[ \rho = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_\omega^2} \]

Since we have \( \sigma_b^2 + \sigma_\omega^2 = \sigma^2 \), we obtain
\[ \sigma_b^2 = \rho \sigma^2 \]

and
\[ \sigma_\omega^2 = (1-\rho) \sigma^2 \]

Now the minimax risk \( r \) can be expressed as
\[
r = \frac{c \sigma^2}{n} + \frac{(1-\rho) \sigma^2}{mn} + c_{1n}m + c_2 n
\]
\[
= \frac{\sigma^2}{mn} [1 + (m-1) \rho] + c_{1n}m + c_2 n
\]

and the optimum values for \( m \) and \( n \) will be
\[
n = \sqrt{\frac{\rho}{c_2 \sigma^2}} \sigma \quad , \quad m = \sqrt{\frac{c_2 1-\rho}{c_1 \rho}}
\]

showing that the optimum intensity of sampling within a cluster depends entirely on the intraclass correlation, and a larger \( \rho \) will require a smaller optimum value for \( m \).

B. Finite populations.

1. Formulation of the problem.

Imagine that nature or some conscious being carried out a fixed sample size experiment, obtaining \((u_1, u_2, \ldots, u_N)\), having chosen some \( \omega \) for the probability density. Later an experiment corresponding to each \( u_i \) (\( i = 1, \ldots, N \)), was carried out, each being of fixed size \( M \), with probability density \( \phi_i \) and such that the first moment of each \( \phi_i \) was equal to \( u_i \). The result obtained is a point \( X \) with \( NM \) coordinates, the coordinate \( X_{ij} \) (\( i = 1, \ldots, N; j = 1, \ldots, M \)) denoting the \( j \)th observation in the experiment with probability density \( \phi_i \) (with mean \( u_i \)). The statistician is told nothing about \( \omega \) of \( \phi_i \) except their variances \( \sigma_b^2 \) and \( \sigma_\omega^2 \), the latter being the same for all \( \phi_i \) and that the mean value associated with \( \phi_i \) is \( u_i \), obtained earlier as a result of an
experiment with probability density \( \omega \). He is given a loss function which does not depend on the form of the distributions \( \omega \) or \( \phi \) and is asked to estimate \( \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij} \). He is, of course, allowed to choose some predetermined number of coordinates of \( X \) and observe their values, and if he prefers, from any number of the sets \( x_i \), where \( x_i = (x_{i1}, \ldots, x_{iM}) \).

The best strategy for the statistician, according to the principles of invariance and sufficiency mentioned earlier in Section III would now be the one which chooses a simple random sample of size \( n \) from the first subscripts, which are altogether \( N \) in number and then corresponding to each subscript so obtained, to choose a sample of \( m \) second subscripts. The coordinates chosen by the statistician to observe will be \( x_{ij} \), \( i = 1, \ldots, n; j = 1, \ldots, m \). The loss function is

\[
L(\bar{u}, f) = (f - \bar{u})^2 + c_1 m + c_2 n
\]

where

\[
\bar{u} = \frac{1}{N M} \sum_{i=1}^{N} \sum_{j=1}^{M} x_{ij} = \frac{1}{N} \sum_{i=1}^{N} u_i
\]

\( c_1 \) = cost of sampling per coordinate \( x_{ij} \)

\( c_2 \) = cost of sampling per set \( (x_i) \), only to determine which \( i \)'s should he sample for \( x_{ij} \), at the second stage.

2. Bayes estimate.

Imagine that somehow the numbers \( n \) and \( m \) have been determined. Now assume that the distribution \( \omega \) which is responsible for \( u_i \)'s is normal with unknown mean \( u \) and known variance \( \sigma_u^2 \). Further, assume that each \( \phi \) which is responsible for \( x_{ij} \)'s is normal with unknown mean \( u \) and known variance \( \sigma_\omega^2 \).

The sample available to the statistician for basing his decision consists of only
\[ X = (X_{11}, \ldots, X_{1m}; X_{21}, \ldots, X_{2m}; \ldots; X_{n1}, \ldots, X_{nm}) \]

and even though he sampled from all \( u_i \)'s \((i = 1, \ldots, N)\), getting \( u = (u_1, \ldots, u_n) \), he was not allowed to observe any \( u_i \). Some partial information available to him concerning \( u_i \) is whatever is provided by \((X_{i1}, \ldots, X_{im})\).

The distribution of \( u_1, \ldots, u_n \), given that \( \sum_{i=1}^{N} u_i = Nu \), where \( u \) is fixed, is \( n \)-variate normal with mean \( u \) for each component and variance covariance matrix

\[
A = \begin{bmatrix} a_{ij} \end{bmatrix} \quad (i, j = 1, \ldots, n)
\]

where

\[
a_{ij} = -\frac{1}{N} \sigma_b^2 \quad \text{for} \quad i \neq j
\]

\[
a_{ij} = \frac{N-1}{N} \sigma_b^2 \quad \text{for} \quad i = j
\]

The distribution of \( X_{i1}, \ldots, X_{im} \), given that \( \sum_{j=1}^{M} X_{ij} = Mu_i \) for each \( i \), is \( m \)-variate normal, with a possibly different mean row \((u_i, u_i, \ldots, u_i)\), but same variance covariance matrix

\[
B = \begin{bmatrix} b_{ij} \end{bmatrix} \quad (i, j = 1, \ldots, m)
\]

where

\[
b_{ij} = -\frac{1}{M} \sigma_\omega^2 \quad \text{for} \quad i \neq j
\]

\[
b_{ij} = \frac{M-1}{M} \sigma_\omega^2 \quad \text{for} \quad i = j
\]

Our first problem is to determine the marginal density of \( X \)'s, from the joint probability density of the sample with \( mn+n \) components, \( n \) being the number of components of \( u \), sampled but unobserved by him and \( mn \) being the number of components for \( X \). This joint probability density of \( X \) and \( u \) can be written as the \( mn+n \) variate normal density:

\[
p(X, u | u) = \text{const.} \ e^{-\frac{1}{2} [(u-u)^' A^{-1} (u-u) + \sum_{i=1}^{n} (X_{i1}-u_{i1})' B^{-1} (X_{i1}-u_{i1})]}\]

where
\[
\bar{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}, \quad \bar{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_m \end{bmatrix}
\]
\[
\bar{x}_l = \begin{bmatrix} x_{l1} \\ \vdots \\ x_{lm} \end{bmatrix}, \quad \bar{x}_l = \begin{bmatrix} u_1 \\ \vdots \\ u_l \end{bmatrix}
\]

and A and B are the \(nxn\) and \(mxm\) matrices defined already.

To integrate it directly over \(u_1, u_2, \ldots, u_n\), is rather complicated. We shall obtain the desired marginal distribution indirectly in the following way.

Let

\[
\bar{z} = \begin{bmatrix} \bar{z}_1 \\ \vdots \\ \bar{z}_{mn} \end{bmatrix}
\]

be the \(mn\times 1\) column vector as the mean and \(C\) an \(mn\times mn\) matrix as the variance covariance matrix of the distribution \(p(x|\bar{u})\). We know that \(\bar{u}\) is the mean and A the covariance matrix in the distribution \(p(u|\bar{u})\).

From the distribution \(p(x|\bar{u})\) and \(p(u|\bar{u})\) we know how to calculate the conditional distribution \(p(x|u,\bar{u})\). For this let us write the covariance matrix of the joint distribution \(p(x,u|\bar{u})\) as

\[
D = \begin{pmatrix} C & E \\ E & A \end{pmatrix}
\]

where \(E\) is due to the fact that the distributions of \(x\) and \(u\) are not independent.

By the well known methods in multivariate analysis, we know now that the mean column vector and the covariance matrix of the conditional distribution \(p(x|u,\bar{u})\) are

\[
\bar{z} + EA^{-1}(u - \bar{u}) \quad \text{and} \quad C - EA^{-1}E \quad \text{respectively.}
\]

Equating these to the corresponding parameters of the distribution \(p(x|u,\bar{u})\), which is
\[ p(x|u, \bar{u}) = \text{const.} e^{-\frac{1}{2} \sum_{i=1}^{n} (x_i - u_i)'^T B^{-1} (x_i - u_i)} \]

we obtain the equations

\[ Z + EA^{-1}(u-u) = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \]

and

\[ C - EA^{-1}E = \begin{pmatrix} B & 0 \\ B & \ddots \\ \vdots & \ddots & B \\ 0 & \cdots & \cdots & B \end{pmatrix} \]

Solving these two equations for \( Z \) and \( C \), we obtain, after some algebraic manipulations,

\[ Z = \begin{bmatrix} \bar{u} \\ \vdots \\ \bar{u} \end{bmatrix} , \]

i.e., each element in \( Z \) is \( \bar{u} \), there being \( mn \) elements in all, and

\[ C = \begin{pmatrix} B + a_{11}F & a_{12}F & \cdots & a_{1n}F \\ a_{21}F & B + a_{22}F & \cdots & a_{2n}F \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1}F & a_{n2}F & \cdots & B + a_{nn}F \end{pmatrix} \]

where \( F \) is an \( mxm \) matrix each of whose elements is unity. Thus we get

\[ p(x|u) = \text{const.} e^{-\frac{1}{2} (x - Z)'C^{-1}(x - Z)} \]

where

\[ x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \]
Now let \( \bar{u} \) have a priori distribution as
\[
P(\bar{u}) = \text{const. } e^{-\frac{1}{2} \frac{\bar{u}^2}{g^2}}
\]
so that we now get
\[
P(\bar{u}|x) = g(x) e^{-\frac{1}{2} (x - \bar{x})' C^{-1} (x - \bar{x}) + \frac{\bar{u}^2}{g^2}}
\]
The expression in the square braces is a quadratic in \( \bar{u} \).

Coefficient of \( \bar{u}^2 = \frac{1}{g^2} + \text{sum of all elements of } C^{-1} \)

Coefficient of \( -2\bar{u} = \sum_{j=1}^{mn} \sum_{i=1}^{mn} x_i \alpha_{ij} \)

where \( \alpha_{ij} \) is the \( i,j \)th element of \( C^{-1} \). Now \( C \) in fact can be written as
\[
G = \begin{pmatrix}
G & H & \cdots & H \\
H & G & H & \cdots \\
: & : & : & \vdots \\
H & \cdots & \cdots & G
\end{pmatrix}
\]
where \( G \) is the matrix with elements
\[
g_{ij} = \frac{N-1}{N} \sigma^2_b + \frac{M-1}{M} \sigma^2_\omega \quad \text{for } i=j
\]
\[
g_{ij} = \frac{N-1}{N} \sigma^2_b - \frac{1}{M} \sigma^2_\omega \quad \text{for } i \neq j
\]
and \( i,j = 1, \ldots, m \), while \( H \) is the matrix with elements
\[
h_{ij} = -\frac{1}{N} \sigma^2_b \quad \text{for all } i,j = 1, \ldots, m
\]
It would be seen that \( C^{-1} \) is also of similar structure, namely, a matrix in all the elements in the diagonal and another matrix everywhere else. The algebraic manipulations finally lead to the following expressions for \( p(\bar{u}|x) \):
\[-\frac{1}{2} \left( \alpha + \frac{1}{\theta^2} \right) (\bar{u} - \frac{\alpha X}{\theta^2})^2 \]

\[p(\bar{u}|X) = k(X) \cdot \theta^{\frac{1}{2}} \]

where

\[\alpha = \frac{1}{\frac{N-n}{N} \sigma_b^2 + \frac{M-m}{M} \sigma_\omega^2} \cdot \frac{n \cdot \text{min}}{m \cdot \text{min}} \]

Hence the Bayes estimate for \( \bar{u} \) is given by

\[\theta^* = \frac{\bar{X}}{1 + \frac{1}{\theta^2} \left[ \frac{1}{n} \left( \frac{1}{N} \right) \sigma_b^2 + \frac{1}{n} \left( \frac{1}{m} \right) \sigma_\omega^2 \right]}^{\cdot} \]

and the Bayes risk is

\[\rho = \frac{\left( \frac{1}{n} - \frac{1}{N} \right) \sigma_b^2 + \frac{1}{n} \left( \frac{1}{m} - \frac{1}{M} \right) \sigma_\omega^2}{\left[ \frac{1}{n} \left( \frac{1}{N} \right) \sigma_b^2 + \frac{1}{n} \left( \frac{1}{m} \right) \sigma_\omega^2 \right]} \cdot \frac{1}{\theta^2} + 1 \]


Now we apply Theorem 4.2 and find that since \( \rho = \frac{1}{\theta^2} \sigma_b^2 + \frac{1}{n} \left( \frac{1}{m} - \frac{1}{M} \right) \sigma_\omega^2 \) as \( \theta \to \infty \), an estimate \( f \) having the same risk as \( r \) will be a minimax estimate. We shall try for the limiting Bayes estimate, which is \( \bar{X} \), the simple mean of all observations.

Moreover, while obtaining risk for estimate \( f = \bar{X} \), we shall not assume normality of the distribution of either \( u \)'s or \( X \)'s. If \( R \) be the risk, we have

\[R = E(\bar{X} \cdot - \bar{u})^2 + c_1 \cdot \text{min} + c_2 \cdot n \]

But \( E(\bar{X} \cdot - \bar{u})^2 \) after simplification, as before, turns out to be equal to

\[\frac{1}{n} \left[ \frac{N-(M-m)}{MM} \sigma_\omega^2 + \frac{N-n}{N} \sigma_b^2 \right] \]

Thus

\[R = \frac{N-n}{Nn} \sigma_b^2 + \frac{M-m}{Mm} \sigma_\omega^2 \cdot \frac{1}{n} + c_1 \cdot \text{min} + c_2 \cdot n \]

\[= r \]
Thus utilizing Theorem 4.3 we find that \( \bar{X} \) is a minimax estimate having a constant risk \( r \), given above, whatever be the underlying distribution which gave rise to the finite populations \( u_1, \ldots, u_N \) and \( x_{ij}, \ldots, x_{iM} \) for each \( i = 1, \ldots, N \).

The approximate optimum values (for minimax sampling) of \( n \) and \( m \) can be obtained as before by differentiating with respect to \( n \) and \( m \) and equating to zero and solving the equation for \( n \) and \( m \). The integers nearest to the values obtained should be approximately correct values.

Proceeding to do that, we notice that we have to solve the same set of equations as in the infinite population case and so we get the same optimum values for \( n \) and \( m \) as before.

4. **Minimax risk in terms of intraclass correlation.**

Just as the intraclass correlation was defined for the infinite populations case, so we define it for finite populations as the simple correlation coefficient between two observations in the same class, whatever has been the original conceptual distribution which gave rise to this finite population. We also denote by

\[
\rho = \frac{1}{NM-1} \sum_{i=1}^{N} \sum_{j=1}^{M} (x_{ij} - \bar{u})^2
\]

which is the variance in case the whole finite population of the size \( NM \) is regarded as having arisen from a single conceptual distribution. The values of \( \rho \) in terms of \( \sigma_b^2 \), \( \sigma_\omega^2 \) and \( \sigma^2 \) is obtained as (for details of derivation, see Sukhatme \(^1\)):

\[
\rho = \frac{M(N-1) \sigma_b^2 - N \sigma_\omega^2}{(NM-1) \sigma^2}
\]

The relation between $\sigma^2$ and $\sigma_b^2$ and $\sigma_\omega^2$ is given by

$$(NM-1) \sigma^2 = (N-1)M \sigma_b^2 + N(M-1) \sigma_\omega^2.$$  

Solving these two equations, we get

$$\sigma_b^2 = \frac{NM-1}{M(N-1)} \frac{\sigma^2}{M} [1 + (M-1)\rho]$$

$$\sigma_\omega^2 = \frac{NM-1}{NM} \sigma^2(1-\rho).$$

The expression for the minimax risk $r$ in terms of $\sigma^2$ and $\rho$ is not a neat looking expression. It is very simple for the cluster sampling, the special case when $m=M$, which we discuss next. It may be of interest, however, to point out that as found before for infinite case, the optimum value for $m$ depends only on $\rho$ and is given by

$$m = \sqrt{\frac{c_2}{c_1} \frac{M(N-1)}{N(M-1)} \frac{1-\rho}{M^{-1} + \rho}}.$$

C. Cluster sampling.

The cluster sampling as pointed out earlier is the special case of two stage sampling discussed above when $m=M$. Conceptually, it is the case where the statistician chooses to observe all the coordinates corresponding to the sampled first subscripts. Considered from this point of view, it becomes just a simple random sample of $u$'s and does not present any theoretical difficulties. In practical applications, however, it is very interesting since it presents a number of problems connected with the size of the first stage units that one should adopt in practice, and we shall not discuss them here. We merely point out that for equal size clusters, of size $M$ units each, the minimax estimate is still a simple mean of all observations, viz., $\bar{X}$, and the minimax risk is obtained by putting $m=M$ in $r$ above. We get
\[ R = \frac{N-n}{N} \sigma_b^2 + c_1 n M + c_2 n \]
\[ = \frac{N-n}{N} \frac{N M - 1}{M(N-1)} \frac{\sigma^2}{M} \left[ 1 + (M-1) \rho \right] + c_1 n M + c_2 n \]

If we agree that \( \frac{N M - 1}{M(N-1)} \) can be approximated to unity without much error, we get
\[ R = \frac{N-n}{N} \frac{\sigma^2}{M} \left[ 1 + (M-1) \rho \right] + c_1 n M + c_2 n \]

The first part, which is the risk due to the errors in estimate, is the variance in the case of cluster sampling, when \( n \) clusters out of \( N \) are chosen, and in this form it was first developed by Hansen and Hurwitz. \( ^1 \)

VIII. Weighted estimate as a minimax procedure.

A. Infinite population.

Suppose there are \( k \) random variables \( X_i \) \( (i = 1, \ldots, k) \) each having the same mean but possibly different variances \( \sigma_i^2 \). A random sample of size \( n_i \) is taken from the \( i \)th random variable and we want to find the minimax estimate, given the loss function
\[ L(u,f) = (f-u)^2 + \sum c_i n_i \]

The problem is exactly similar to the one discussed in stratified sampling excepting that there the means in the different strata were possibly different and we were required to estimate \( u = \sum N_i u_i \), some linear function of the various \( u_i \)'s.

1. Bayes procedures.

Again we see that there cannot be a least favorable distribution, and we obtain a sequence of Bayes procedures by taking the a priori distribution of \( u \) as

p(u) = const. \( e^{\frac{-1}{2} \frac{u^2}{\sigma^2}} \)

Note that sampling is done independently for each \( X_i \), and \( E(X_i) = u \) for all \( i \). The mean of the observations obtained for any \( X_i \) will thus be a sufficient estimate for \( u \) and so we may consider \( f \) as a function of \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \). To get a Bayes estimate for given \( n_i \), we proceed exactly as in the previous cases, by assuming normal distribution for each \( X_i \), and get

\[
-\frac{1}{2} \sum_{i=1}^{k} \frac{n_i}{\sigma_i^2} (\bar{x}_i - u)^2
\]

\[p(X|u) = \text{const.} \ e^{\frac{-1}{2} \frac{u^2}{\sigma^2}}\]

where \( X \) denotes \( \bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k \). Since

\[
-\frac{1}{2} \frac{u^2}{\sigma^2}
\]

\[p(u) = \text{const.} \ e^{\frac{-1}{2} \frac{u^2}{\sigma^2}}\]

we obtain

\[-\frac{1}{2} \left[ \sum \frac{n_i}{\sigma_i^2} (\bar{x}_i - u)^2 + \frac{u^2}{\sigma^2} \right] \]

\[p(u|X) = c(X) \cdot e^{-\frac{1}{2} \sum \frac{n_i}{\sigma_i^2} (\bar{x}_i - u)^2 - u^2} \sum \frac{n_i}{\sigma_i^2} \]

\[= c(X) \cdot e^{-\frac{1}{2} \sum \frac{n_i}{\sigma_i^2}} \left[ (u - f_\theta)^2 \right] \]

where

\[f_\theta = E(u|X) = \frac{\sum \frac{n_i \bar{x}_i}{\sigma_i^2}}{\sum \frac{n_i}{\sigma_i^2} + \frac{1}{\sigma^2}} \]

is the Bayes estimate, and the Bayes risk is

\[r_\theta = \sum \frac{n_i}{\sigma_i^2} + \frac{1}{\sigma^2} \sum \sigma_i n_i \]
2. Minimax procedures.

The minimax procedure would be obtained by utilizing Theorem 4.2. We see that as \( \theta \to \infty \)

\[
    r_\theta \to r = \frac{1}{\sum \frac{n_i}{\sigma_i^2}}
\]

and any estimate having this risk will be a minimax estimate.

Let us try the estimate obtained as limit of a Bayes estimate when \( \theta \to \infty \). This is the so-called weighted estimate, which we can write down as

\[
    f = \frac{\sum \omega_i x_i}{\sum \omega_i}
\]

where

\[
    \omega_i = \frac{n_i}{\sigma_i^2}
\]

In order to show that \( f \) is minimax we have to find its risk and show that it is equal to \( r \). We note that \( f \) is a linear function of the normally distributed and independent chance variables \( X_i \). They are known to have variance as \( \frac{\sigma_i^2}{n_i} = \frac{1}{\omega_i} \). The risk corresponding to \( f \) is then seen to be

\[
    R = E(f-u)^2 + \sum c_i n_i
\]

\[
    = \frac{\sum \omega_i^2}{(\sum \omega_i)^2} + \sum c_i n_i = \frac{1}{\sum \omega_i} + \sum c_i n_i
\]

\[= r\]

Hence the estimate \( f \) is minimax, when we know that \( X \)'s are distributed normally.

Now with the help of Theorem 4.3 we can generalize this statement to the case when we know nothing about the distribution of \( X \)'s, save that they all have the same mean and variances in the different class, (i.e., variance of the random variables \( X_i \)) are \( \sigma_i^2, i=1,2,...,k \).
The risk corresponding to the minimax estimate found above, will then be, for given \( n_i \),

\[
R = E(f-u)^2 + \sum c_i n_i \\
= E\left(\frac{\sum \omega_i \bar{X}_i - u \sum \omega_i}{\sum \omega_i}\right)^2 + \sum c_i n_i \\
= \frac{1}{\left(\sum \omega_i\right)^2} E\left[ \sum \omega_i^2 (\bar{X}_i - u)^2 + \sum_i \sum_{i \neq j} \omega_i \omega_j (\bar{X}_i - u)(\bar{X}_j - u) \right] + \sum c_i n_i \\
= \frac{1}{\left(\sum \omega_i\right)^2} \sum \omega_i^2 \sigma_i^2 + \sum c_i n_i \\
= \frac{1}{\sum \omega_i} + \sum c_i n_i = r.
\]

Since the risk is constant and equal to the minimax risk found earlier for normality of \( X \)'s, we conclude that the weighted mean

\[
f = \frac{\sum \omega_i \bar{X}_i}{\sum \omega_i}, \quad \omega_i = \frac{\sigma_i^2}{n_i}
\]

is a minimax estimate of the mean \( u \) whatever be the distributions of \( X_i \), so long as all \( X_i \) have the same mean \( u \) and variances \( \sigma_i^2 \). It would be noticed that the weights are the reciprocal of the variances.

The optimum values for \( n_i \) can be chosen by minimizing \( r \) by the usual methods and obtaining the integral values of the set \( (n_1, \ldots, n_k) \) for which \( r \) will be least.

B. Finite populations.

Suppose now that there are \( k \) finite populations \( (X_{i1}, X_{i2}, \ldots, X_{iN_i}) \) where \( i = 1, 2, \ldots, k \) as a result of \( k \) experiments conducted by nature with some distributions of which we know nothing except that each has the same mean and it is given that the variance of the \( i^{th} \) distribution is \( \sigma_i^2 \). The statistician is faced with the problem of estimating the common mean of the distributions used.
in these experiments and is allowed to observe a sample of observations from each of the k populations. He is given the cost function

$$L(u, r) = (f - u)^2 + \sum c_i n_i$$

where $c_i$ is the cost of taking observations from the $i$th population.

If $k$ were only one, the best strategy, insofar as sampling is concerned, the discussion in Section III above would show that, as a result of the principle of invariance and sufficiency, the statistician should choose a sample by the simple random sampling method. Now instead of one, there are several populations and suppose he takes a random sample from each one of them to estimate the mean of each population. But he knows that all these estimates are actually the estimates of the same quantity he was asked to estimate, viz., the common mean $u$. He, therefore, would like to know if he can get a better estimate than the individual estimates, and also in so doing, considering the risks involved in the various estimates (which include the corresponding sampling costs too), how best to decide on $n_i$ so as to obtain an overall estimate which has minimax risk.

Let us first consider the problem of estimation for some given values of $n_i$. Once the minimax risk is found for given $n_i$, he can choose optimum values of $n_i$ simply by minimizing that risk over the various values of $n_i$.

Since no least favorable distribution exists for the same reasons as before, we find Bayes estimates assuming first that the distribution that gave rise to the $k$ populations was normal in each case.

Noticing that the sample mean is a sufficient statistic for estimating the population mean, we can consider estimates as functions of the means $\bar{X}_1, \ldots, \bar{X}_k$. Since the conditional distribution of mean $\bar{X}_i$, given $\sum_{j=1}^{k} X_{ij} - \bar{X}_i = \text{const.}$ is normal with mean $u$ and variance $\frac{N_i - n_i}{N_i} \frac{\sigma_i^2}{n_i}$, we obtain
\[-\frac{1}{2} \sum_{i=1}^{k} \frac{N_i n_i}{N_i - n_i} \frac{(\bar{X}_i - u)^2}{\sigma_i^2}\]

where \(X\) stands as before for \((\bar{X}_1, \ldots, \bar{X}_k)\). Now if we let \(\frac{N_i}{N_i - n_i} \sigma_i^2 = v_i\), the whole process goes through as in the case of infinite population with \(v_i\) in place of \(\sigma_i^2\), and we get the Bayes estimate

\[
f_\theta = \frac{\sum n_i \bar{X}_i}{\sum n_i + \frac{1}{\theta^2}}
\]

\[
= \sum n_i \bar{X}_i \frac{1}{N_i - n_i} \sigma_i^2 + \frac{1}{\theta^2}
\]

and the Bayes risk

\[
r_\theta = \frac{1}{\sum n_i \bar{X}_i \frac{1}{N_i - n_i} \sigma_i^2 + \frac{1}{\theta^2} + \sum c_i n_i}
\]

As \(\theta \to \infty\),

\[
r_\theta \to r = \frac{1}{\sum n_i \bar{X}_i \frac{1}{N_i - n_i} \sigma_i^2 + \sum c_i n_i}
\]

Now we shall show that an estimate (which is the Bayes limiting estimate as \(\theta \to \infty\)) obtained by weighting the means \(\bar{X}_i\) with weight inversely proportional to the variance \(\frac{N_i}{N_i - n_i} \sigma_i^2\) is a minimax estimate. This limiting Bayes estimate is

\[
f = \lim_{\theta \to \infty} f_\theta = \frac{\sum \omega_i \bar{X}_i}{\sum \omega_i}
\]
where
\[ \omega_i = \frac{N_i n_i}{N_i - n_i} \frac{1}{\sigma_i^2} . \]

Let us find the risk with this estimate, without assuming any underlying conceptual distribution for \( X \)'s. If \( R \) be the risk, we have

\[
R = E(f-u)^2 + \sum c_i n_i \\
= \frac{1}{(\sum \omega_i)^2} \left[ \sum \omega_i^2 (\bar{x}_i - u)^2 + \sum \omega_i \sum \omega_j (\bar{x}_i - u)(\bar{x}_j - u) \right] + \sum c_i n_i \\
= \frac{1}{(\sum \omega_i)^2} \sum \omega_i^2 \frac{N_i - n_i}{N_i n_i} \sigma_i^2 + \sum c_i n_i \\
= \frac{1}{\sum \omega_i} + \sum c_i n_i = r .
\]

The risk is constant and equal to the limiting value \( r \) of the sequence of Bayes risks. By utilizing Theorem 4.2 and 4.3, we conclude that, since we have not assumed any particular distribution for \( X \)'s in deriving the risk for \( f \) above, the weighted estimate \( f \), the weights being inverse of the variance of the means, is a minimax estimate. The optimum values for \( n_i \) can be chosen as usual.

**IX. Ratio estimate as a minimax procedure.**

In the last section we found under what circumstances a weighted mean, with weights inversely proportional to the variances, turned out to be a minimax estimate. In this section we shall consider under what circumstances we could obtain "ratio estimate" as a minimax estimate.

Ratio estimate arises in sample surveys in a natural way when sampling is done on units which are of unequal sizes. Consider a character \( Y \) (standing
for size, in general) which is already known accurately for each member of
the population consisting of \( N \) units or can be obtained exactly for every unit
in the sample, and suppose a random variable \( X \) is observed for each sampling
unit chosen. If \( X_1, \ldots, X_n \) be the observations on \( X \) and \( Y_1, \ldots, Y_n \) be the
exact values of \( Y \), the so-called ratio estimate for estimating the total of
the population, \( \sum_{i=1}^{N} X_i \), is
\[
\frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i} \cdot \frac{\sum_{i=1}^{N} Y_i}{N}
\]

It is called ratio estimate since we are using \( \sum_{i=1}^{n} X_i \) as the estimate for
the population ratio \( \frac{\sum_{i=1}^{N} X_i}{\sum_{i=1}^{N} Y_i} \). We shall show that under certain conditions
this is a minimax estimate.

A. Infinite population.

Suppose that with a random variable \( Y \), with known mean and any distribution
whatsoever, is associated another random variable \( X \) in the following way.
Nothing is known about the distribution of \( X \) except that the conditional
distribution of \( X \), given \( Y = y_i \), say, has both its mean and variance proportional
to \( y_i \). There are available \( n \) observations on \( Y \) and corresponding to each
\( Y = y_i \) \( (i = 1, \ldots, n) \) we have an observation \( X = x_i \) on \( X \). We want to find a
minimax estimate for \( E(X) \), with the loss function \( L(u, f) = (f-u)^2 + \alpha \).

Specifically, suppose that
\[
E(X|Y = y_i) = \alpha y_i \quad \chi \text{ unknown}
\]
\[
\text{Var}(X|Y = y_i) = \beta y_i \quad \beta \text{ known}
\]
We can define now a new random variable for each fixed \( Y = y_1 \) \( (y_1 \neq 0) \) as
\[
Z_i = \frac{X}{y_i}
\]
which has the property that each \( Z_i \) has a common mean \( \alpha \) and the variance (may be different) of \( Z_i \) is \( \frac{\beta}{y_i} \).

This is now reduced to exactly the previous case where the weighted mean was found to be minimax, for estimating the common mean. Here the common mean of \( Z_i \)'s is \( \alpha \). The weights to be used are the reciprocal of the known variances of the means of \( Z_i \)'s. In the present case there is only one observation for each \( Z_i \) and so the observed values \( Z_i \) are to be weighted with
\[
\omega_i = \frac{1}{\sigma_i^2} = \frac{y_i}{\beta}.
\]
Thus we obtain the minimax estimate for estimating \( \alpha \) as
\[
f_Z = \frac{\sum_{i=1}^{n} \frac{y_i X_i}{\beta y_i}}{\sum_{i=1}^{n} \frac{y_i}{\beta}} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i}.
\]

Now we know that
\[
E(X) = E\left\{E(X|Y = Y_1)\right\} = \alpha E(Y) = \alpha E(Y).
\]

Since \( E(Y) \) is known and the minimax estimate for \( \alpha \), i.e., of \( E(X)/E(Y) \) is \( f_Z \), we obtain, the minimax estimate of \( E(X) \) as
\[
f = f_Z \cdot E(Y) = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i} E(Y)
\]
which is the ratio estimate.

Since we have already fixed each \( n_i = 1 \), we have not considered the sampling costs while deriving this estimate.

It might be of interest to point out that we have used each \( n_i = 1 \) in order to show how the minimax estimate turns out to be simply the usual ratio estimate, where the ratio is taken of all the sample values of \( X \) to the
sample values of $Y$. It may be argued, however, that since the sampling of $X$ has been done independently for each $Y = y_i$, one might like to have more observations on $X$ for a given $Y$ and not just one. The case still satisfies the conditions where minimax estimate turns out to be weighted estimate. If we, therefore, take $n_i$ instead of one as the number of observations on $X$ for a given $Y = y_i$, $(i = 1, \ldots, k)$ and $\sum n_i = n$, then minimax estimate for $\alpha$ would be as before

$$f_Z = \frac{\sum_{i=1}^{k} n_i y_i}{\sum_{i=1}^{k} n_i y_i}$$

sample total of $X$ values

total of $Y$ values, one for each $X$ in the sample

The only modification introduced is that we have to include for each observation $X$ the corresponding value of $Y$ in the denominator total. The estimate is still called the ratio estimate, and minimax estimate of $E(X)$ is given by

$$f = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} Y_i} = \frac{E(X)}{E(Y)}$$

Since the risk, apart from the cost of experimentation, in the minimax estimate when it turns out to be weighted mean is

$$\frac{1}{\sum \omega_i} = \frac{1}{\sum_{i=1}^{k} \frac{n_i y_i}{\beta}}$$

we obtain minimax risk, for given $n_i$, for the estimate of $E(X)$ as

$$r = \frac{\beta}{\sum_{i=1}^{k} n_i y_i} \frac{[E(Y)]^2}{\sum_{i=1}^{k} c_i n_i}$$
In the above, $\beta$, $Y_i$, $c_i$, $(i = 1, \ldots, k)$, and $E(Y)$ are all given and so one can determine the optimum values of $n_i$ (for minimax sampling) by minimizing the total minimax risk.

B. Finite populations.

Consider a finite population $X_1, \ldots, X_N$, which is divided into $k$ classes with the help of an auxiliary characteristic $Y$, known for each unit in the population, there being $N_i$ units in the $i$th class, each unit having $Y = Y_i$ $(i = 1, \ldots, k)$. The mean and variance of the finite population in the $i$th class (having known number $N_i$ units) are proportional to $Y_i$ each, the constant of proportionality $(= \beta)$ known for variance, but $(= \alpha)$ unknown for mean. A minimax estimate is required for the total of the $X$'s in the entire population, with loss function being

$$L(u, f) = (f - u)^2 + \sum_{i=1}^{k} \frac{c_i n_i}{N_i}$$

where $f$ is the estimate of $u = \sum_{i=1}^{k} \sum_{j=1}^{N_i} X_{ij}$, and $c_i$ the cost of experimentation per unit in the $i$th class.

Suppose somehow the statistician has decided on the number $n_i$ of $X$'s he is going to sample from the $i$th class. Define for each class another characteristic as the ratio of $X$ to the corresponding $Y$, viz., $Z_i = X_i / Y_i$. Since $Y_i$ is fixed for the $i$th class, and since the mean of all $X_{ij}$ in that class $(j = 1, \ldots, N_i)$ is $\alpha Y_i$, the mean of all $Z_{ij}$ in the $i$th class becomes the unknown $\alpha$ (common to all the $k$ classes), while since the variance of $X_{ij}$ in the $i$th class is given to be $\beta Y_i$, the variance of $Z$'s in the $i$th class becomes $\beta / Y_i$, a known value since both $\beta$ and $Y_i$ are known. The $Z$'s now satisfy the conditions in the last section under which a minimax estimate for the common mean was found to be a weighted mean of $\bar{Z}_i$'s, the means of $Z$'s in the various classes, the weights used being the inverse of the variances of the means. The minimax estimate for $\alpha$ is, therefore,
\[ f_Z = \frac{\sum_{i=1}^{k} \omega_i \bar{Z}_i}{\sum_{i=1}^{k} \omega_i} \]

where

\[ \omega_i = \frac{N_i n_i}{(N_i - n_i) \sigma_i^2} \]

\[ = \frac{N_i n_i}{(N_i - n_i) \beta} \]

But \( \bar{Z}_i = \bar{X}_i / Y_i \), and so by substitution, we get

\[ f_Z = \frac{\sum_{i=1}^{k} \frac{N_i n_i}{(N_i - n_i) \beta} \bar{X}_i / Y_i}{\sum_{i=1}^{k} \frac{N_i n_i}{(N_i - n_i) \beta} Y_i} \]

\[ = \frac{\sum_{i=1}^{k} \frac{N_i n_i}{N_i - n_i} \bar{X}_i}{\sum_{i=1}^{k} \frac{N_i n_i}{N_i - n_i} Y_i} \]

Now if \( n_i \) were such that \( n_i \ll N_i \), then

\[ \frac{N_i}{N_i - n_i} = \frac{N}{N-n} \]

and we obtain

\[ f_Z = \frac{\sum_{i=1}^{k} n_i \bar{X}_i}{\sum_{i=1}^{k} n_i Y_i} \]

\[ = \frac{\text{sum of all } X_i's \text{ in the sample}}{\text{sum of } Y_i's \text{ corresponding to each } X \text{ in the sample}} \]

\( f_Z \) is thus the ratio of the total of \( X_i's \) in the sample to the total of corresponding \( Y_i's \) and is a minimax estimate for \( \alpha \), which is the ratio of the total of all \( X_i's \) in the entire population to the total of all \( Y_i's \) in the whole population as seen below:
\[
\sum_{j=1}^{N_i} x_{ij} = N_i (\alpha Y_i)
\]

Adding over all classes, we get
\[
\sum_{i=1}^{k} \sum_{j=1}^{N_i} x_{ij} = \alpha \sum_{i=1}^{k} N_i Y_i
\]
or
\[
\alpha = \frac{\text{total of all } X_i's}{\text{total of all } Y_i's} = \frac{u}{T}
\]

where \( T \) denotes the known total of all \( Y_i \)'s.

Hence, the minimax estimate of \( u = T \alpha \) is given by
\[
f = T z
\]

= ratio in the sample multiplied by the total of all \( Y_i \)'s in the population.

The minimax risk is given by
\[
r = \frac{T^2}{N-n} \sum_{l=1}^{k} n_i Y_i / \beta + \sum c_i n_i
\]
since \( n_i \) has been assumed to be proportional to \( N_i \) in the above derivation.

We can now choose the optimum values of \( n_i \) by minimizing \( r \) over possible values of \( n_i \). Let
\[
\frac{n_i}{N_i} = q = \frac{n}{N}
\]
then we get
\[
r = \frac{(1-q)T^2}{N-n} \alpha + q \sum_{l=1}^{k} c_i N_i
\]
\[
= T \beta \frac{1-q}{q} + q \mathcal{C}
\]
where \( \mathcal{C} = \text{cost of complete enumeration} \). To minimize \( r \), we have
\[
\frac{\partial x}{\partial q} = -\frac{T}{q^2} + c = 0
\]
or
\[
q = \sqrt{\frac{Tc}{c}}
\]
where \(T, \beta,\) and \(C\) are all known.

The value of \(q\) thus obtained may not give integral values for \(n_1 = qN_i\)
but the approximation to the nearest integral value should be very satisfactory.

As pointed out earlier, if we took the loss function as
\[
L(u, f) = a(f - u)^2 + \sum c_i n_i
\]
we should get all the results by simply changing \(c_i\) into \(c_i/a\), and multiplying
risk by \(a\). The minimax estimate is exactly the same, while the minimax risk
is now
\[
R = aT \beta \left( \frac{1 - q}{q} \right) + q C
\]
and the optimum value of the sampling fraction \(n/N = q\) is given by
\[
q = \sqrt{\frac{aT\beta}{C}}
\]

It would be of interest to note here that Cochran¹ proved that for
large \(N\) (also \(N_i\) large for each \(i\)) if
(i) the relation between \(X\) and \(Y\) is a straight line through the
origin, and
(ii) the variance of \(X\) about this line is proportional to \(Y\),
then the ratio estimate is "best unbiased linear estimate".

¹W. G. Cochran, "Sampling theory when the sampling units are of unequal
X. **Regression estimate as a minimax procedure.**

It should be noticed that in the derivation of the ratio estimate as a minimax estimate we did not consider \( X \) and \( Y \) together as a random variable, neither did we assume anything about the variance of \( Y \), nor about the possible correlation between \( X \) and \( Y \). In fact, we shall see in this section that if \( X \) and \( Y \) be taken together as random variables and it is assumed that their variances, correlation, and mean of \( Y \) are known, then the minimax estimate turns out to be the usual regression estimate.

A. **Infinite population.**

Let \( X \) and \( Y \) be distributed jointly in a bivariate normal distribution, the distribution known exactly but for the mean of \( X \), say \( u \). A random sample \( (X_1,Y_1;\ldots;X_n,Y_n) \) of size \( n \) is obtained from this distribution. We are interested in getting a minimax estimate for \( u \), with loss function as

\[
L(u,f) = (f-u)^2 + cn
\]

where \( c \) is the total cost of experimentation per sampling unit (consisting of cost on \( X \) and \( Y \) both), while \( f \) is the proposed estimate for \( u \).

1. **Bayes estimate.**

We observe again that there is no least favorable distribution and we would find a sequence of Bayes estimates against the sequence of a priori distributions of \( u \), as

\[
p(u) = \text{const.}, \quad e^{-\frac{1}{2} \frac{u^2}{\sigma^2}}
\]

The procedure will be exactly the same as before, and the Bayes estimate (since we have quadratic loss function) will be the mean of \( u \) in the conditional distribution of \( u \) given the sample \( (X_1,Y_1;\ldots;X_n,Y_n) \).

The probability density of the bivariate normal distribution is given by

\[
\text{const.} \cdot e^{-\frac{1}{2} \left[ \frac{(x-u)^2}{\sigma_x^2} + \frac{2\rho(x-u)(y-v)}{\sigma_x\sigma_y} + \frac{(y-v)^2}{\sigma_y^2} \right]}
\]

The only unknown parameter is \( u \), the remaining parameters \( v, \sigma_x, \sigma_y, \rho \).
are assumed to be known.

For simplicity, we shall let \( \frac{Y - \bar{Y}}{\sigma_Y} = z \), then if \( x \) stands for \( (x_1, \ldots, x_n) \) and \( z \) for \( (z_1, \ldots, z_n) \) for short, we can write the joint probability density of the sample \( (x, z) \) as

\[
p(x, z|u) = \text{const.} e^{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left[ \frac{(X_i - u)^2}{\sigma^2} - \frac{2 \rho (X_i - u) Z_i}{\sigma} + Z_i^2 \right] - \frac{u^2}{2\theta^2}}
\]

where we are writing \( \sigma^2 \) for \( \sigma_x^2 \). Thus we get

\[
p(u|x, z) = c(x, z) e^{-\frac{1}{2(1-\rho^2)} \sum_{i=1}^{n} \left[ \frac{(X_i - u)^2}{\sigma^2} - \frac{2 \rho (X_i - u) Z_i}{\sigma} + Z_i^2 \right] - \frac{u^2}{2\theta^2}}
\]

\[
= c'(x, z) e^{-\frac{1}{2(1-\rho^2)} \left[ u^2 \left( \frac{\theta^2}{\sigma^2} + \frac{1-\rho^2}{\theta^2} \right) - 2u \left( \frac{\sum X_i}{\sigma^2} - \frac{\rho \sum Z_i}{\sigma} \right) \right]}
\]

\[
= c''(x, z) e^{-\frac{1}{2(1-\rho^2)} \left( \frac{n}{\sigma^2} + \frac{1-\rho^2}{\theta^2} \right) (u - f_\theta)^2}
\]

where

\[
f_\theta = \frac{\sum X_i - \rho \sigma \sum Z_i}{n \sigma^2 (1-\rho^2) / \theta^2 + 1 + \sigma^2 (1-\rho^2) / n \theta^2}
\]

As before, \( f_\theta \) is Bayes estimate for given \( n \), and the Bayes risk is given by

\[
r_\theta = \frac{1-\rho^2}{\sigma^2 + \frac{1-\rho^2}{\theta^2}} + cn
\]

2. **Minimax estimate.**

Let \( \theta \to \infty \), then \( r_\theta \to r = \frac{\sigma^2 (1-\rho^2)}{n} + cn \). Now by Theorem 4.2, if there is an estimate \( f \), which has its risk constant and equal to \( r \), then \( f \) is a minimax estimate.
Let us try, as a possible estimate, the limiting Bayes estimate as \( \theta \to \infty \).

It is given by

\[
f = \lim_{\theta \to \infty} f_{\theta} = \bar{X} - \rho \bar{\sigma} \bar{Z}
\]

\[
= \frac{\rho \bar{\sigma}}{\bar{\sigma}_y} \left( \bar{y} - v \right)
\]

by substituting back the values of \( \bar{\sigma} \) and \( \bar{Z} \).

Since we would eventually like to remove the assumption of normality of the bivariate distribution of \( X \) and \( Y \), let us try to find the risk corresponding to \( f \), without assuming anything about the joint distribution of \( X \) and \( Y \) excepting that their variances are \( \sigma_x^2 \) and \( \sigma_y^2 \), the correlation is \( \rho \) and mean of \( Y \) is \( v \). The risk \( R \) under these conditions is given by

\[
R = E(f-u)^2 + cn
\]

\[
= E[(x - u) - \frac{\rho \sigma_x}{\sigma_y} (y - v)]^2 + cn
\]

\[
= E[(x - u)^2 - \frac{2 \rho \sigma_x}{\sigma_y} (x - u)(y - v) + \frac{\rho^2 \sigma_x^2}{\sigma_y^2} (y - v)^2] + cn
\]

\[
= \frac{1}{n} \left[ \sigma_x^2 - \frac{2 \rho \sigma_x}{\sigma_y} \rho \sigma_y + \frac{\rho^2 \sigma_x^2}{\sigma_y^2} \right] + cn
\]

\[
= \frac{\sigma_x^2}{n} (1 - \rho^2) + cn = r
\]

Thus, utilizing Theorem 4.3 we conclude that \( f \) is a minimax estimate, whatever be the joint distribution of \( X \) and \( Y \) excepting that the variances of \( X \) and \( Y \), the correlation \( \rho \) and mean of \( Y \) be known.

We now proceed to determine the optimum value for \( n \) (for minimax sampling) by minimizing the minimax risk \( r \) over all possible values of \( n \). Let \( r(n) \) stand for value of \( r \) when a sample of \( n \) is chosen from the bivariate distribution.
\[ r(n) = \frac{\sigma^2}{n} (1 - \rho^2) + cn \]

where \( \sigma^2 \) is the variance of \( X \)

\[ r(n+1) = \frac{\sigma^2}{n+1} (1 - \rho^2) + c(n+1) \]

Therefore

\[ r(n+1) - r(n) = c = \frac{\sigma^2(1 - \rho^2)}{n(n+1)} \]

The optimum value of \( n \) is obtained by choosing the first integer for

which the above difference changes sign from negative to positive, i.e., the

value for which the inequality

\[ (n + \frac{1}{2})^2 \geq \frac{\sigma^2(1 - \rho^2)}{c} + \frac{1}{4} \]

holds good for the first time. Thus

\[ n = \text{integer nearest to } \sqrt{\frac{\sigma^2(1 - \rho^2)}{c} + \frac{1}{4}} \]

and when \( n \) is exactly between two integers, say \( m + \frac{1}{2} \), it does not matter

which one of the nearest integers \( m \) or \( m+1 \) is taken for \( n \).

B. Finite populations.

The formulation of the problem of sampling from a finite population

(Blackwell and Girshick\(^1\)) mentioned earlier in Section III applies equally

well whether each \( X_i \) \((i = 1, \ldots, N)\) of the finite population is a real number

or a vector with real components. Thus the principles of invariance and

sufficiency lead us to a strategy of simple random sampling from the finite

population \((X_1, Y_1; \ldots; X_N, Y_N)\), i.e., sampling \( n \) out of \( N \) vectors, each with

a probability of \( \frac{1}{N} \). It is proposed to estimate \( u \), the mean of \( X \),

assuming known the mean of \( Y \), the variances of \( X \) and \( Y \) and their correlation,

adopting the loss function \( L \) as

\[ L(u, f) = (f - u)^2 + cn \]

\(^1\)Blackwell, David and Girshick, M. A., ibid.
1. Bayes estimates.

As before, there does not exist a least favorable distribution for \( u \). We shall, therefore, follow the same method of finding a sequence of Bayes estimates, as before. Also we shall assume that the conceptual distribution which gave rise to the finite population was bivariate normal, with \( u \) and \( v \) as means of \( X \) and \( Y \), and

\[
\begin{pmatrix}
\sigma_x^2 & \rho \sigma_x \sigma_y \\
\rho \sigma_x \sigma_y & \sigma_y^2
\end{pmatrix}
\]

as the variance-covariance matrix.

The conditional distribution of a sample \( (X_1, Y_1; X_2, Y_2; \ldots; X_n, Y_n) \) from a bivariate normal distribution, given \( X_1^+ \ldots + X_N = \text{const.} \), \( Y_1^+ \ldots + Y_N = \text{const.} \), is a \( 2n \)-variate normal distribution with covariance matrix as

\[
G = \begin{pmatrix}
A & B & \ldots & B \\
B & . & & . \\
\vdots & \vdots & & \vdots \\
B & \ldots & B & A
\end{pmatrix}
\]

where \( A \) and \( B \) are \( 2 \times 2 \) covariance matrices,

\[
A = \left(1 - \frac{1}{N}\right) \begin{pmatrix}
\sigma_x^2 & \rho \sigma_x \sigma_y \\
\rho \sigma_x \sigma_y & \sigma_y^2
\end{pmatrix}
\]

\[
B = -\frac{1}{N} \begin{pmatrix}
\sigma_x^2 & \rho \sigma_x \sigma_y \\
\rho \sigma_x \sigma_y & \sigma_y^2
\end{pmatrix}
\].

Let us take a priori density of \( u \) to be normal with mean zero and variance \( \sigma^2 \), i.e.,

\[
p(u) = \text{const.} \ e^{-\frac{1}{2} \frac{u^2}{\sigma^2}}
\]
The probability density of the sample \((x, y)\) where \((x, y)\) stands for the sample \((x_1, y_1; x_2, y_2; \ldots; x_n, y_n)\) is 2n-variate normal, given by

\[
p(x, y \mid u) = \text{const. } e^{-\frac{1}{2} Q},
\]

where

\[
Q = \begin{pmatrix}
  X - u \\
  Y - v \\
  \vdots \\
  X - u \\
  Y - v
\end{pmatrix}
\]

\[
Q^{-1} = \begin{pmatrix}
  X - u \\
  Y - v \\
  \vdots \\
  X - u \\
  Y - v
\end{pmatrix}
\]

therefore

\[
p(u \mid x, y) = c(x, y) e^{-\frac{1}{2} (Q + \frac{u^2}{\sigma^2})}
\]

Now \(Q + \frac{u^2}{\sigma^2}\) is a quadratic expression in \(u\), and we want to express it as a square of a linear form in \(u\). Thus we first find the coefficients of \(u^2\) and \(u\) in \(Q\).

By symmetry

\[
Q^{-1} \equiv \begin{pmatrix}
  A & B & \ldots & B \\
  B & \ddots & \ldots & \vdots \\
  \vdots & \ddots & B \\
  B & \ldots & B & A
\end{pmatrix}^{-1} = \begin{pmatrix}
  C & D & \ldots & D \\
  D & \ddots & \vdots & \vdots \\
  \vdots & \ddots & D \\
  D & \ldots & D & C
\end{pmatrix}
\]

where

\[
AC + (n-1)BD = I
\]

\[
AD + BC + (n-2)BD = 0
\]

However,

\[
A = (1 - \frac{1}{N}) \sum
\]

\[
B = - \frac{1}{N} \sum
\]

where
\[ \sum = \begin{pmatrix} \sigma_x^2 & \rho \sigma_x \sigma_y \\ \rho \sigma_x \sigma_y & \sigma_y^2 \end{pmatrix} \]

Thus, we get
\[ \sum \left[ \frac{-N-1}{N} C - \frac{n-1}{N} D \right] = I \]
\[ \sum \left[ -\frac{1}{N} C + \frac{N-n+1}{N} D \right] = 0 \]

These equations reduce to
\[ (N-1)C - (n-1)D = N \sum^{-1} \]
\[ -C + (N-n+1)D = 0 \]
giving
\[ [(N-1)(N-n+1) - (n-1)D = N \sum^{-1} \]
or
\[ D = \frac{1}{N-n} \sum^{-1} \]

and, therefore,
\[ C = \frac{N-n+1}{N-n} \sum^{-1} \]

Thus, if we let
\[ C = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \]

and
\[ D = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \]

we find that
\[ d_{11} = \frac{1}{(N-n)(1-\rho^2)} \frac{1}{\sigma_x^2} \]
\[ d_{12} = d_{21} = -\frac{1}{(N-n)(1-\rho^2)} \frac{\rho}{\sigma_x \sigma_y} \]
\[ d_{22} = \frac{1}{(N-n)(1-\rho^2)} \frac{1}{\sigma_y^2} \]

and

\[ c_{ij} = (N-n+1)d_{ij} \quad \text{for } i=1,2; \; j=1,2. \]

Now we can write \( Q \) as

\[
Q = \begin{pmatrix} X_1-u \\ Y_1-v \\ \vdots \\ X_n-u \\ Y_n-v \end{pmatrix} \begin{pmatrix} G & D & \cdots & D \\ D & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & D \\ D & \cdots & D & C \end{pmatrix} \begin{pmatrix} X_1-u \\ Y_1-v \\ \vdots \\ X_n-u \\ Y_n-v \end{pmatrix}
\]

\[
= c_{11} \sum_{i=1}^{n} (X_i-u)^2 + 2c_{12} \sum_{i=1}^{n} (X_i-u)(Y_i-v) + c_{22} \sum_{i=1}^{n} (Y_i-v)^2 \\
+ d_{11} \sum_{i \neq j}^{n} (X_i-u)(X_j-u) + 2d_{12} \sum_{i \neq j}^{n} (X_i-u)(Y_j-v) + d_{22} \sum_{i \neq j}^{n} (Y_i-v)(Y_j-v).
\]

Therefore the coefficient of \( u^2 \) in \( Q \)

\[
= nc_{11} + n(n-1)d_{11} = Nnd_{11}.
\]

Coefficient of \( u \) in \( Q \)

\[
= -2c_{11} \sum_{i=1}^{n} X_i - 2c_{12} \sum_{i=1}^{n} (Y_i-v) - d_{11} \sum_{i \neq j}^{n} (X_i+X_j) - 2(n-1)d_{12} \sum_{i=1}^{n} (Y_j-v) \\
= -2\left[ c_{11} + (n-1)d_{11} \right] \sum_{i=1}^{n} X_i + \left[ c_{12} + (n-1)d_{12} \right] \sum_{i=1}^{n} (Y_i-v) \\
= -2\left[ Nnd_{11} \bar{X} + Nnd_{12} (\bar{Y}-v) \right].
\]

Thus
\[- \frac{1}{2} (Q + \frac{u^2}{e^2}) \]
\[p(u|x,y) = c(x,y) \cdot e^{\frac{1}{2} [u^2 (N_n d_{11} + \frac{1}{e^2}) - 2u (N_n d_{11} \bar{x} + N_n d_{12} \bar{y})]} = c'(x,y) \cdot e^{-\frac{1}{2} (N_n d_{11} + \frac{1}{e^2}) (u - f_\theta)^2} = c''(x,y) \cdot e \]

where
\[f_\theta = \frac{N_n [d_{11} \bar{x} + d_{12} \bar{y}]}{N_n d_{11} + \frac{1}{e^2}}.\]

As before, the Bayes estimate is \(E(u|x,y) = f_\theta\), and the Bayes risk is \(r_\theta = \frac{1}{N_n d_{11} + \frac{1}{e^2}} + cn\), the values of \(d_{11}\) and \(d_{12}\) being given above.


We see that as \(\theta \to \infty\),
\[r_\theta \to r = \frac{N_n d_{11}}{N_n} \int (1 - \rho^2) + cn\]

Thus, if we can find any estimate having its risk equal to \(r\), that estimate will be minimax.

Such an estimate is provided by the limiting Bayes estimate
\[f = \lim_{\theta \to \infty} f_\theta = \bar{x} + \frac{d_{12}}{d_{11}} (\bar{y} - \bar{x}) = \bar{x} - \frac{\rho \sigma x}{\sigma y} (\bar{y} - \bar{x})\]
as we shall show presently.

If \(R\) be the risk corresponding to \(f\), we obtain, without assuming any underlying conceptual distribution,
\[ R = \mathbb{E}(f - u)^2 + cn \]
\[ = \mathbb{E}[(x - u - \rho \frac{\sigma_x}{\sigma_y} (y - v))^2] + cn \]
\[ = \mathbb{E}(x - u)^2 - 2 \rho \frac{\sigma_x}{\sigma_y} \mathbb{E}(x - u)(y - v) + \rho^2 \frac{\sigma_x^2}{\sigma_y^2} \mathbb{E}(y - v)^2 + cn \]
\[ = \frac{N-n}{Nn} \left[ \sigma_x^2 - 2 \rho^2 \sigma_x^2 + \rho^2 \sigma_x^2 \right] + cn \]
\[ = \frac{N-n}{Nn} \sigma_x^2 (1 - \rho^2)^2 + cn = r \].

Hence, utilizing the Theorem 4.3 we see that \( f \) is a minimax estimate, whatever be the form of the underlying conceptual distribution which could be responsible for the finite population.

We may mention, however, that we shall define correlation in a finite population by
\[ \rho \sigma_x \sigma_y = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - u)(y_i - v) \]

where
\[ u = \frac{1}{N} \sum_{i=1}^{N} x_i \] , \[ v = \frac{1}{N} \sum_{i=1}^{N} y_i \]
\[ \sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - u)^2 \] , \[ \sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - v)^2 \].

Thus
\[ \rho = \frac{\sum (x_i - u)(y_i - v)}{\sqrt{\sum (x_i - u)^2 \cdot \sum (y_i - v)^2}} \].

In defining these parameters in this way, we would see that the covariance matrix for a finite population is defined exactly as the covariance matrix of the conceptual distribution used by nature or some conscious being to obtain the finite population.
As far as the optimum value for $n$ is concerned, it will be seen that $r(n+1) - r(n)$ is exactly the same as for infinite population and so the optimum value for $n$ is again the same, viz.,

$$n = \text{integer nearest to } \frac{\sqrt{\frac{\sigma^2_x (1 - \rho^2)}{c}} + \frac{1}{4}}{}$$
XI

Literature cited.


6. Lehman, E. L., Mimeographed notes on the Theory of Estimation, University of California, Berkeley, Chap. IV.


XII Acknowledgment.

I am deeply indebted to Professor M. A. Girshick, who encouraged me to work on this problem, suggested the topic, and gave generous help and guidance. I am deeply grateful to the Office of Naval Research for financial support. My grateful thanks are also due to Professors David Blackwell, Herman Chernoff, and Herman Rubin for the numerous helpful discussions and suggestions during the course of the entire work.