Bayesian Design of Single and Double Stratified Sampling for Estimating Proportion in Finite Population†

S. Zacks

The University of New Mexico

0. Introduction

The present paper is based on an earlier study of the author [11], and its primary objective is to present the methodology of a Bayesian design of stratified sampling. In order to avoid abstract mathematics, and make the exposition more concrete, we focus attention on a problem of estimating a linear function of the proportion defectives in k finite lots of a certain product. In particular, consider the following sampling inspection problem, k lots (k ≥ 2) of a certain product are subject to sampling inspection. The sizes of the lots, N1, N2, · · · , Nk, say, are known. Let Mi (i = 1, · · · , k) be the number of defective items in the i-th lot. The objective is to estimate the total number of defective items θ = Σi Mi, in the k lots. A total sample of a fixed size, n, is specified. The problem is how to allocate the sample over the k given lots. We study this allocation problem in a Bayesian framework with a squared-error loss. Thus, whatever sampling plan is adopted, the estimator of θ to be used after observing the sample is the Bayes estimator for the squared-error loss. The question is, however, what sampling plan to adopt. The principle one following is that of determining a sampling plan which minimizes the associated prior risk. In a recent paper [12] the author discussed the problem of the optimal Bayes selection of a sample from a finite population. It has been shown there that the optimal Bayes sampling selection is nonrandomized, without replacement, and should often be performed sequentially. However, to obtain such an optimal Bayes sampling plan, it is required to specify a prior joint distribution to all the N variates (X1, · · · , XN) of the population. This requirement is very often impractical. It is rarely the case that the sampling designer can specify such a joint prior distribution. Bayesians tend often to specify a simple joint prior distribution, which can be conveniently manipulated. Under such simplified prior assumptions one is liable to attain trivial designs. A further elaboration of this point can be found in Solomon and Zacks [10]. We feel that most practitioners, especially, in the field of quality control, where they have to inspect

† Partially supported by National Science Foundation Grant GP-9007.
Received Nov. 1967; revised June 1969.
the quality of lots produced by different manufacturers may lack the complete prior information required for a purely Bayesian sampling design. Furthermore, it is a common practice to sample the items from the finite lots in a simple random manner, without replacement. We shall therefore fix attention on such a random sampling, and reduce the whole decision problem to that of determining the optimal allocation.

Bayes and minimax allocations were previously studied by Aggarwal [1, 2], Ericson [5, 6, 7]; and following Zacks [11] we have the extensive study of Grosh [8]. The studies of Aggarwal and Ericson deal only with single-phase stratified sampling. These are designs which are completely specified before observations commence. Optimal Bayes double-phase designs were studied by Zacks [11], Grosh [8] and Draper and Guttman [4]. Zacks [11] considered also sequential designs. We shall not discuss here sequential designs, but will confine attention to the methodology of determining the double-phase optimal Bayes designs. Draper and Guttman [9] assumed the mythical assumption that the within-stratum distribution is normal. We tackle the problem by concentrating on the case of hypergeometric distributions of the number of defectives observed in each lot. These hypergeometric distributions are generated by the simple random sampling without-replacement. We assume that the unknown parameters \( M_1, \ldots, M_k \), which designate the number of defectives in the \( k \) lots, are priorly independent random variables, having uniform distributions on the set of integers \( \{1, 2, \ldots, N_i\}, i = 1, \ldots, k \). This assumption reflects a state of complete ignorance of the quality controller about the number of defectives in the incoming lots. Even if this assumption is too simplified, the controller has an occasion, after analyzing the observations obtained in the first phase of sampling, to adjust and design the allocation of the second phase sample according to the first phase posterior distribution of \( (M_1, \ldots, M_k) \). This posterior distribution depends on the number of defectives in each lot found in the first phase of sampling, and thus yields a prior for the second phase which is closer to reality. Another reason for adopting uniform priors for \( (M_1, \ldots, M_k) \) in the present exposition is that the Bayes second-phase allocation becomes similar to the classical Neyman's allocation in sampling theory (see Raj [9] pp. 65). The sampling designer can, however, proceed along the methodological framework described here to derive Bayes double-phase designs with respect to any prior distribution more appropriate for his situation. We remark that certain prior distributions (like binomial priors) yield posterior risks independent of the observations. In these cases the optimal Bayes designs are single-phase ones (see Zacks [11]). As we shall illustrate numerically, in the case of uniform priors, the double-phase optimal designs do not reduce the prior risk significantly over that of single-phase designs. From a practical point of view, single-phase designs are in many cases more manageable than double-phase designs. Thus, it is interesting to know for which prior distributions the single-phase designs are close in their prior risks to double-phase designs. Grosh [8] studied this question with respect to the wide class of prior beta-binomial distributions. These prior distributions do not yield, in general, a significant gain in optimal double-phase design. It is important to investigate the effect of double-phase designs in cases of other prior distributions.
1. The Statistical Model and the Problem of Optimal Allocation of a Single-Phase Stratified Sample

Given \( k \) strata of \( N_1, N_2, \cdots, N_k \) units, let \( M_1, \cdots, M_k \) be the number of defective units in the strata (lots). From each stratum a simple random sample is drawn without replacements. The \( k \) samples are independent and of size \( n_1, \cdots, n_k \) respectively. Let \( X_i \) (\( i = 1, \cdots, k \)) denote the number of defective units in the \( i \)-th sample. \( X_1, \cdots, X_k \) are independent, and have hypergeometric distributions with densities

\[
    f(x|N_i, M_i, n_i) = \binom{M_i}{x} \binom{N_i - M_i}{n_i - x} / \binom{N_i}{n_i}, \quad \text{if } x = 0, 1, \cdots, n_i
    \]

We are concerned with the estimation of a parameter \( \theta = \sum_{i=1}^{k} x_i \lambda_i P_i \), where \( \lambda_1, \cdots, \lambda_k \) are specified coefficients, and \( P_i = M_i/N_i \) (\( i = 1, \cdots, k \)). It is well known that the best unbiased estimator of \( \theta \) is

\[
    \hat{\theta}_n = \sum_{i=1}^{k} \lambda_i \frac{X_i}{n_i}
    \]

The variance of this estimator is

\[
    \text{Var} \{ \hat{\theta}_n \} = \sum_{i=1}^{k} \lambda_i P_i (1 - P_i) \left( 1 - \frac{n_i - 1}{N_i - 1} \right)
    \]

Suppose that the cost of sampling is \( \$c_i \) (\( i = 1, \cdots, k \)) for observing one unit of the \( i \)-th stratum. Furthermore, let \( C \) be the budget available for sampling. An allocation of the stratified sample is a choice of \( k \) non-negative integers \( (n_1, \cdots, n_k) \) which satisfy the constraint \( C = \sum_{i=1}^{k} c_i n_i \). An allocation will be denoted by \( \mathbf{n} \). An allocation \( \mathbf{n}^* \) is called optimal if it minimizes (1.3). Using the Lagrange multiplier technique, we can easily show that the optimal allocation for (1.3) is given approximately by:

\[
    n_i^* = C \frac{\lambda_i}{\sum_{i=1}^{k} \lambda_i \sqrt{c_i P_i (1 - P_i)}} \frac{1}{\sqrt{N_i - 1}}, \quad i = 1, \cdots, k
    \]

where

\[
    a_i = |\lambda_i| \sqrt{N_i / (N_i - 1)}, \quad i = 1, \cdots, k
    \]

For large values of \( N_i \), \( a_i = |\lambda_i| \) (\( i = 1, \cdots, k \)) measures the relative "importance" of sampling from the \( i \)-th stratum. The problem is that \( \mathbf{n}^* \) depends on the unknown values of \( P_1, \cdots, P_k \), and we generally cannot attain an optimal allocation without further assumptions. The methodology of sampling surveys treats this problem by substituting in (1.4) estimates of \( P_i \) (\( i = 1, \cdots, k \)) available from previous surveys (see Cochran [3], p. 81).
2. **Bayes Optimal Single-Phase Allocation**

Consider the case of priorly independent $M_1, \ldots, M_k$, with uniform prior distributions over $\{0, 1, \ldots, N_i\}$. Let $\xi_i(m)$ designate the prior density function of $M_i$ then, for each $i = 1, \ldots, k$,

\[(2.1) \quad \xi_i(m) = \begin{cases} (N_i + 1)^{-1}, & \text{if } m = 0, 1, \ldots, N_i \\ 0, & \text{otherwise} \end{cases} \]

Using the identity (see Appendix)

\[(2.2) \quad \sum_{x=0}^{n_i} \binom{r+x}{r} \binom{N-x-r}{n-x} = \binom{N+1}{N-n}, \]

we obtain that the marginal density of $X_i$, in which $M_i$ is mixed with the prior (2.1) is, for each $i = 1, \ldots, k$,

\[(2.3) \quad f(x | \xi_i, N_i, n_i) = \frac{1}{N_i+1} \sum_{x=0}^{N_i} \frac{m(N_i-m)}{\binom{N_i}{n_i}} \binom{m}{x} \frac{N_i-m}{n_i-x} \]

\[= \frac{1}{N_i+1} \sum_{x=0}^{N_i} \frac{m(N_i-m)}{\binom{N_i}{n_i}} \binom{m}{x} \frac{N_i-m}{n_i-x} \]

\[= \frac{1}{N_i+1}, \quad x = 0, \ldots, n_i. \]

Furthermore, the posterior density of $M_i$ ($i = 1, \ldots, k$), given $\{X_i = x\}$ is:

\[(2.4) \quad \pi_i(m|N_i, n_i, x) = \frac{\binom{m}{x} \frac{N_i-m}{n_i-x}}{\binom{N_i+1}{N_i-n_i}}, \quad m = x, \ldots, N_i - n_i + x. \]

The Bayes estimator of $P_i = M_i/N_i$ for the squared-error loss, based on the sample from the $i$th stratum ($i = 1, \ldots, k$) is, when $\{X_i = x\}$,

\[(2.5) \quad \hat{P}_i = \frac{1}{N_i(N_i+1)} \sum_{x=0}^{N_i} \binom{m}{x} \frac{N_i-m}{n_i-x} = \frac{x+1}{n_i+2} \left(1 - \frac{2}{N_i}\right) - \frac{1}{N_i}. \]

The Bayes estimator of $\theta = \sum_i P_i$ is therefore:

\[(2.6) \quad \hat{\theta} = \sum_{i=1}^{k} \lambda_i \left[ \frac{X_i + 1}{n_i + 2} \left(1 + \frac{2}{N_i}\right) - \frac{1}{N_i}\right]. \]

In the special case of estimating the total number of defectives $\theta = \sum_i M_i$, we substitute in (2.6), $\lambda_i = N_i (i = 1, \ldots, k)$ and obtain the Bayes estimator

\[\sum_{i=1}^{k} \frac{N_i + 2}{n_i + 2} (X_i + 1) - k.\]
We derive now the prior risk function $R(\xi, n)$ associated with the uniform priors, and a given allocation $n$. Due to the prior independence $\lambda_i, \cdots, \lambda_k$, the posterior risk of the Bayes estimator $\theta$ for a squared-error loss, coincides with the posterior variance of $\theta$, given $(X_1, \cdots, X_k)$. To derive this posterior variance we notice that

$$(2.7) \quad \frac{1}{N + 1} \sum_{x=1}^{N+n} m^x \left( \frac{m}{N + x} \right)$$

$$\quad = (N + 2)(N + 3) \left( \frac{x + 1}{n + 2} \right) - 3(N + 2) \frac{x + 1}{n + 2} + 1.$$

Thus, we obtain

$$(2.8) \quad \text{Var} \{ \theta \mid n_i, X_1, \cdots, X_k \}$$

$$\quad = \sum_{i=1}^{k} \lambda_i^2 \frac{(X_i + 1)(n_i + X_i)}{(n_i + 2)(n_i + 3)} \frac{N_i + 2}{N_i} \left( 1 - \frac{n_i}{N_i} \right).$$

The expectation of (2.9) with respect to the marginal uniform densities specified by (2.3) is obtained in the following manner. Since,

$$(2.10) \quad \frac{1}{n + 1} \sum_{x=1}^{n} m^x \left( \frac{m}{n + x} \right) = \frac{1}{6(n + 2)}$$

we obtain the prior* Bayes risk:

$$(2.11) \quad R(\xi, n) = \frac{1}{6} \sum_{i=1}^{k} \lambda_i^2 \frac{N_i + 2}{N_i} \left( 1 - \frac{n_i}{N_i} \right).$$

The optimal allocation of the sample is found by the common Lagrangian method in the following manner. Letting

$$(2.12) \quad \rho_i = \left( C + 2 \sum_{i=1}^{k} \lambda_i \right) \left( \frac{1 + 2}{N_i} \right) \sqrt{\epsilon_i}$$

we define

$$(2.13) \quad \bar{\lambda}_i = (\rho_i - 2)^*, \quad i = 1, \cdots, k,$$

where $\alpha^* = \max (0, \alpha)$. If some of the $\bar{\lambda}_i$ values are zero we reallocate the sampling budget $C$ to the strata for which $\bar{\lambda}_i > 0$. For this, let $A$ denote a subset of

* It is a common fashion to call $R(\xi, n)$ the preposterior risk. However, $R(\xi, n)$ is exactly the minimal attainable prior risk. We reserve the term preposterior risk for the case of double stratified sampling, in which the preposterior risk is the conditional expectation of the posterior risk, given the observations of the first phase of sampling.
for which \( n_i > 0 \). Let,

\[
(2.14) \quad n_i^0 = \begin{cases} 
0, & \text{if } i \notin A \\
(\rho_i^* - 2)^+, & \text{if } i \in A,
\end{cases}
\]

where, for each \( i \in A \)

\[
(2.15) \quad \rho_i^* = (C + 2 \sum_{i \in A} \sum_{j=1}^{N_i} e_i) \frac{|\lambda_i| \left(1 + \frac{2}{N_i}\right)}{\sum_{i \in A} |\lambda_i| \left(1 + \frac{2}{N_i}\right)} \sqrt{e_i}.
\]

If, for some \( i \in A \), \( (\rho_i^* - 2)^+ = 0 \) we set \( n_i = 0 \) for that \( i \), remove \( i \) from \( A \), and reallocate \( C \) over the strata whose indices are in \( A \). This process continues until \( (\rho_i^* - 2)^+ > 0 \) for all \( i \in A \). Since this allocation procedure yields non-integer values for \( n_i^0 \), we shall rule, for simplicity that the actual sample size \( n_i^0 \) is the least integer greater or equal to the value obtained from (2.14) and (2.15). The minimal Bayes prior risk \( R(\xi, n^0) \) is obtained by substituting the optimal allocation values \( n_i^0 (i = 1, \cdots, k) \) in (2.11). A numerical example is given in section 4.

3. BAYES OPTIMAL DOUBLE-PHASE ALLOCATION

In the double stratified sampling the sampling is performed in two phases. In the first phase a budget of \( C_1 \) is allocated to the \( k \) strata, and in the second phase a budget of \( C_2 \) is allocated. Let \( m \) denote the first phase allocation, and \( n \) the second phase allocation. Let \( X = (X_1, \cdots, X_k) \) denote a vector of independent hypergeometric random variables (number of defectives) observed in the first phase sample and let \( Y = (Y_1, \cdots, Y_k) \) denote the observed vector of the second phase sample. Given \( X, Y_1, \cdots, Y_k \) are conditionally independent having certain hypergeometric distributions. The problem is three fold:

(i) What is the optimal allocation of the total budget \( C \) into the two phases, \( C_1 \) and \( C_2 \), \( C_1 + C_2 = C \)?

(ii) What is the optimal allocation of the first phase sampling, \( m^0 \)?

(iii) What is the optimal allocation of the second phase sampling, \( n^0 \)?

Given the allocation \( (m, n) \) and the observed vectors \( (X, Y) \) the Bayes estimator of \( \theta = x^F \) is

\[
(3.1) \quad \hat{\theta}(X, Y; m, n) = \sum_{i=1}^{k} \lambda_i \left[ \frac{X_i + Y_i + 1}{n_i + m_i + 2} \left(1 + \frac{2}{N_i}\right) - \frac{1}{N_i} \right].
\]

The associated Bayes posterior risk is

\[
(3.2) \quad R(X, Y; m, n) = \sum_{i=1}^{k} \lambda_i \left(1 + \frac{2}{N_i}\right) \frac{X_i + Y_i + 1}{n_i + m_i + 1} \frac{1 - X_i - Y_i}{n_i + m_i + 2} \left(1 - \frac{n_i + m_i}{N_i}\right) \left(1 - \frac{X_i + Y_i}{n_i + m_i + 3}\right) \left(1 - \frac{n_i + m_i}{N_i}\right).
\]

After the first-phase of the sampling we have the allocation \( m \) and the observed vector \( X \). If a second-phase allocation \( n \) is chosen the expected Bayes posterior
risk, which will be called the preposterior risk, is

\begin{equation}
R(X; m, n) = E_{Y \mid X, m, n}\{R(X, Y; m, n)\}.
\end{equation}

\(E_{Y \mid X, m, n}\{\cdot\}\) designates the expectation of the term in brackets with respect to the preposterior distribution of \(Y\). This preposterior distribution is the mixture of the distribution of \(Y\) with respect to the posterior distribution \((M_i, \cdots, M_k)\), given \(X, m, n\). The preposterior distribution of \(Y_i(i = 1, \cdots, k)\) is given, as proven in the Appendix by

\begin{equation}
f_y(y \mid X_i = x, m_i, n_i) = \begin{cases} 
\frac{(n_i)(m_i)}{(x + y)} \frac{m_i + 1}{n_i + m_i + 1}, & y = 0, \cdots, n_i \\
0, & \text{otherwise}.
\end{cases}
\end{equation}

Hence, from (3.2) and (3.4) we obtain the preposterior risk:

\begin{equation}
R(X; m, n) = \sum_{i=1}^{k} \frac{\lambda_i}{N_i} \left(1 + \frac{2}{N_i}\right) \left(1 - \frac{m_i + n_i}{N_i}\right) \cdot \sum_{y=0}^{n_i} f_y(y \mid X_i, m_i, n_i) \frac{(X_i + 1 + y)(m_i + n_i + 1 - X_i - y)}{(m_i + n_i + 2)^2(m_i + n_i + 3)}.
\end{equation}

Straightforward algebraic manipulations yield

\begin{equation}
R(X; m, n) = \sum_{i=1}^{k} \lambda_i \left(1 + \frac{2}{N_i}\right) \left(\frac{X_i + 1)(m_i + 1 - X_i)}{(m_i + 2)(m_i + 3)(m_i + n_i + 2)} \left(1 - \frac{m_i + n_i}{N_i}\right)\right). 
\end{equation}

The second-phase optimal allocation, given \(C, m, X\) is an allocation \(n^0\) which minimizes (3.6). Comparing (2.11) to (3.6) we deduce that the optimal second-phase allocation is determined by the following algorithm:

Start with the set \(A = \{1, 2, \cdots, k\}\) and determine for each \(i \in A\)

\begin{equation}
\rho_i = (C + \sum_{i \in \tilde{A}} c_i) \frac{\lambda_i \left(1 + \frac{2}{N_i}\right)}{\sum_{i \in \tilde{A}} \lambda_i \left(1 + \frac{2}{N_i}\right)} \sqrt[4]{\frac{(X_i + 1)(m_i + 1 - X_i)}{(m_i + 2)(m_i + 3)} \left(1 - \frac{m_i + n_i}{N_i}\right)}.
\end{equation}

Let \(\tilde{A} = (\rho_i - m_i - 2)^{+}\). If for some \(i \in A\), \(\tilde{A}_i = 0\) set \(n_i^0 = 0\) and delete that index from \(A\). Determine again the values of \(\rho_i\) and \(\tilde{A}_i\) for \(i \in \tilde{A}\). If for some indices of \(A\) \(\tilde{A}_i = 0\) set \(n_i^0 = 0\) and delete these indices from \(A\). Repeat this process until \(\tilde{A}_i > 0\) for all \(i \in A\). Then set \(n_i^0 = \text{least integer} \geq \tilde{A}_i (i \in \tilde{A})\). We remark that the optimal Bayes allocation of the second-phase sampling is a function of \(X, m, n^0 = n^0(X, m)\). Thus, the sample size from each stratum, at the second-phase of sampling, is a random variable. If we substitute \(n^0(X, m)\) in (3.6) we obtain the expected Bayes risk, after the first-phase sampling, for the optimal second-phase allocation, which is
To determine the optimal first-phase allocation \( m^0 \), we have to determine first the expectation of (3.8) with respect to the uniform prior marginal distributions (2.3) of \( X_i \). It is not a simple matter to determine this prior expectation of the first term on the R.H.S. of (3.8). The prior expectation of the second term there is \( \frac{1}{2} \sum_i \lambda_i (1 + 2/N_i) \). One could try various approximations to the prior expectation of (3.8). Instead, we shall examine numerically in the next section a pseudo-optimal determination of the first-phase allocation, namely the allocation of the first-phase sample according to the Bayes optimal single-phase allocation, as described in section 2.

4. A Numerical Example

In the present numerical example we consider the following simple special case:

(i) The number of strata: \( k = 2 \).
(ii) The size of strata: \( N_1 = N_2 = 500 \).
(iii) Cost of observation: \( c_1 = c_2 = 1 \) [$.]

Letting \( U_i = (X_i (1 + 2/N_i), i = 1, 2 \) and a total sample of size \( n + m = 100 \) (i.e., \( C = 100 \) [$.]) we calculate numerically all possible optimal second-phase allocation, as \( X_i = 0, 1, \cdots, m_i (i = 1, 2) \). These values are then substituted in (3.6) to yield the associated Bayes posterior risks. These Bayes posterior risks are then averaged with respect to the prior uniform marginal distributions of \( X_1 \) and \( X_2 \). These computations have been performed for all possible first-phase allocation. The allocations yielding the minimal prior risks, and the values of these prior risks are presented in Table 1. In this table we take various values of \( m \) (total first sample size); fix \( a_1 = 1.0 \) and vary \( a_2 \) over the range \( [1.0, 1.5, 2.0, 3.0, 5.0, 10.0] \). The values in the column of \( m = 100 \) yield the optimal Bayes single-phase allocation and the associated prior risk values.

Examining the prior risk values presented in Table 1 we arrive at the following conclusions:

(i) For each value of \( a_2 \), the variation in the minimal prior risk values as a function of \( m \) is very slow. All the differences for double-phase sampling are in the fourth decimal place.
(ii) There is not much difference between the prior risks of the optimal single-
Table 1*

Optimal First-Phase Allocations and Prior Bayes Risks \( \times 10^5 \) [8] in Allocations of Double-Phase Stratified Sampling

<table>
<thead>
<tr>
<th>( m )</th>
<th>( n_0 )</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
<th>80</th>
<th>90</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
<td>( n_1 )</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>5.656</td>
<td>5.656</td>
<td>20</td>
<td>2.534</td>
<td>2.578</td>
<td>30</td>
<td>2.578</td>
<td>2.578</td>
</tr>
<tr>
<td>1.5</td>
<td>10</td>
<td>8.772</td>
<td>8.772</td>
<td>20</td>
<td>8.704</td>
<td>8.686</td>
<td>23</td>
<td>8.677</td>
<td>8.677</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>43.313</td>
<td>43.313</td>
<td>10</td>
<td>40.307</td>
<td>40.307</td>
<td>10</td>
<td>40.307</td>
<td>40.307</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>159.400</td>
<td>159.400</td>
<td>5</td>
<td>159.116</td>
<td>159.116</td>
<td>5</td>
<td>159.070</td>
<td>159.070</td>
</tr>
</tbody>
</table>

* The values of this table were computed by Dr. D. L. Grosh (see [8]).

The optimal first-phase total sample size increases with \( \alpha_2 \).

(iii) The optimal first-phase total sample size increases with \( \alpha_2 \).

As mentioned in section 3, since it requires a considerable amount of computations to determine the optimal first-phase allocation, in a double sampling procedure, we tried a pseudo-optimal Bayes allocation. In this allocation, the first-phase sample is allocated according to the optimal Bayes single-phase procedure, as given in section 2. The second-phase sample is allocated in an optimal Bayes manner, as given in section 3. The values of the first-phase sample allocation and the associated prior risk values are given in Table 2.

The comparison of the prior risk values of the pseudo-optimal procedure (Table 2) to those of the optimal procedure (Table 1) reveals that, although the first phase allocations are different, the differences in the prior risk values are very small. The conclusion is that, unless we are aiming at very high precisions (in which differences in the fourth and fifth decimal places are significant), the pseudo-optimal Bayes allocation of a double-phase stratified sampling is very close to optimal. Taking into account the simplicity of the pseudo-optimal procedure relative to the optimal procedure, there is a good reason to recommend the use of the pseudo-optimal allocation procedure.
TABLE 2

Pseudo-Optimal First-Phase Allocations and Prior Risks \( \times 10^4 \) in a Double-Phase Stratified Sampling

<table>
<thead>
<tr>
<th>( m )</th>
<th>20</th>
<th>30</th>
<th>40</th>
<th>50</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>( m_1 )</td>
<td>( R )</td>
<td>( m_1 )</td>
<td>( R )</td>
<td>( m_1 )</td>
</tr>
<tr>
<td>1</td>
<td>10</td>
<td>5.636</td>
<td>15</td>
<td>5.608</td>
<td>20</td>
</tr>
<tr>
<td>1.5</td>
<td>9</td>
<td>3.777</td>
<td>12</td>
<td>3.733</td>
<td>16</td>
</tr>
<tr>
<td>5</td>
<td>3.1</td>
<td>49.649</td>
<td>4</td>
<td>49.575</td>
<td>5</td>
</tr>
<tr>
<td>10</td>
<td>0.61</td>
<td>159.61</td>
<td>1</td>
<td>159.50</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( m )</th>
<th>60</th>
<th>70</th>
<th>80</th>
<th>90</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n )</td>
<td>( m_1 )</td>
<td>( R )</td>
<td>( m_1 )</td>
<td>( R )</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>5.571</td>
<td>35</td>
<td>5.575</td>
</tr>
<tr>
<td>1.5</td>
<td>26</td>
<td>8.677</td>
<td>28</td>
<td>8.678</td>
</tr>
<tr>
<td>2</td>
<td>21</td>
<td>12.121</td>
<td>23</td>
<td>12.127</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>49.301</td>
<td>10</td>
<td>49.800</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>159.12</td>
<td>5</td>
<td>159.04</td>
</tr>
</tbody>
</table>

* The values of this table were computed by Dr. D. L. Grosh (see [8]).

5. Appendix

In the present Appendix we prove identity (2.2) and verify (3.4). We start with the proof of (2.2).

Consider the identity

\[
\sum_{j=0}^{\infty} \binom{m}{j} P^j (1 - P)^{m-j} = 1, \quad \text{for all } 0 < P < 1.
\]

Mixing the binomial terms of (5.1) with a beta distribution we obtain

\[
\frac{1}{B(v + 1, q + 1)} \sum_{j=0}^{\infty} \binom{m}{j} \int_0^1 P^j (1 - P)^{m-j} dP = 1,
\]

for any \( 0 < v, q < \infty \). In particular for \( v \) and \( j \) integers we can write (5.2) in the following form

\[
\sum_{j=0}^{\infty} \binom{j + v}{j} \binom{m - j + q}{q} / \binom{m + v + q + 1}{m} = 1.
\]

Substituting in (5.3) \( m = N - n, q = n - x, v = x \) we obtain (2.2). To prove (3.4) we start from
Bayesian Stratified Sampling for Estimating a Proportion

\[ f(y \mid X_i = x, m_i, n_i) = \sum_{x=y}^{n_i-m_i+x} \left( \binom{M}{x} \binom{N_i - M}{m_i - x} \binom{M - x}{n_i - y} \right) \]

where the function \( \psi(N_i, m_i + n_i, x + y) \) is determined in the following manner: First, since \( \sum_{x=0}^{n_i} P(y \mid X_i = x, m_i, n_i) = 1 \) for all \( (x, m_i, n_i), i = 1, \ldots, k \), we obtain from (5.1) the identity

\[ \sum_{x=0}^{n_i} \binom{N_i}{x} \binom{m_i + 1}{n_i + 1} \psi(N_i, m_i + n_i, x + y) = 1. \]

On the other hand, starting from the identity \( \sum_{x=0}^{n_i} P(x \mid P) = 1 \) for all \( 0 < P < 1 \), we obtain by mixing it with a beta distribution for \( P \),

\[ \sum_{x=0}^{n_i} \binom{N_i}{x} \frac{1}{B(x + 1, m_i + 1)} \int_0^1 P(x \mid P) \frac{1}{m_i + n_i + 1} \frac{1}{x + y} dP = 1. \]

Comparing (5.2) and (5.3) we obtain, since these identities hold for all \( (N_i, m_i + n_i, x + y) \), that

\[ \psi(N_i, m_i + m, x + y) = \frac{N_i + 1}{n_i + m_i - 1} \frac{1}{m_i + n_i} \frac{1}{x + y}, \quad i = 1, \ldots, k. \]

Finally, substituting (5.7) in (5.4) we obtain (3.4).

References

