A Double Sampling Scheme for Estimating from Misclassified Multinomial Data with Applications to Sampling Inspection

AARON TENENBEIN
Department of Quantitative Analysis
New York University
New York, New York

In some situations, it is desired to estimate multinomial proportions from data which have been misclassified. One such area is the sampling inspection area of quality control. In this paper, it is assumed that two measuring devices are available to classify units into one of r mutually exclusive categories. The first device is an expensive procedure which classifies units correctly; the second device is a cheaper procedure which tends to misclassify units. In order to estimate the proportions \( p_i \) (\( i = 1, 2, \ldots, r \)) a double sampling scheme is presented. At the first stage, a sample of \( N \) units is taken and the fallible classifications are obtained; at the second stage a subsample of \( n \) units is drawn from the main sample and the true classifications are obtained. The maximum likelihood estimates of the \( p_i \) are derived along with their asymptotic variances. Optimum values of \( n \) and \( N \) which minimize the measurement costs for a fixed precision of estimation and which minimize the precision for fixed cost are derived. The procedures of estimation and allocation of sample size are illustrated by an example in quality control.

KEY WORDS
Double Sampling Scheme
Misclassified Data
Estimation
Sample Size Determination
Sampling Inspection
Minimize Variance
Minimize Cost

1. INTRODUCTION

The effects of misclassification on the sample estimate of the binomial proportion has been studied by Bross [1] under the assumption of a probabilistic misclassification model. He has shown that, under misclassification, the sample proportion is a biased estimate of the binomial proportion, \( p \), and that the bias, which is a function of the amount of misclassification in the data, can be substantial. Bross’s work on the effects of misclassification can be extended to multinomial data with similar results; namely, the sample estimates \( \hat{p}_i \) \( (i = 1, 2, \ldots, r) \) are biased estimates of the multinomial proportions \( p_i \), and these biases can be substantial.

Received Nov. 1969; revised Aug. 1970

187
In order to adjust for this bias, some knowledge of the extent of misclassification which is present in the data must be available. One method of obtaining information on the extent of misclassification is to compare the results obtained by two or more measuring devices on the same group of sampling units. Let us assume that an investigator has available two measuring devices or classifiers. Suppose further that the first classifier is a relatively inexpensive procedure which is subject to misclassification error, and that the second classifier is a more expensive procedure which is subject to no misclassification error. Thus the second measuring device is considered to be the true classifier as opposed to the former fallible classifier.

This situation may occur in the sampling inspection area of quality control. Here, the fallible classifier may be the inspector whereas the preferred true classification procedure would be a more expensive procedure to apply in practice. One example of such a procedure may be destructive sampling, that is, sampling which would result in the destruction of the unit under inspection.

The problem facing the investigator is as follows: unbiased estimates of the multinomial proportions can be obtained by the true classifier; however, this may prove to be very expensive. On the other hand, if the fallible classifier is used, the sample proportions may be biased, perhaps very seriously. In this paper a compromise between these two extremes is discussed, namely a double sampling approach. First, a random sample of \( N \) units is drawn and the fallible classification is obtained for each member in the sample. Then, from the \( N \) units already drawn, a random sample of \( n \) units \( (n < N) \) is drawn and the true classification is obtained for each of these \( n \) units. Thus, there are a total of \( n \) units in the sample which have been classified by both the true and fallible devices. The data on these \( n \) units provide some information regarding the extent of misclassification which is present. The multinomial proportions can thus be estimated from the available data without going to the extreme of obtaining the true classification for all \( N \) units in the sample.

Tenenbein [3] used this double sampling scheme to estimate from misclassified binomial data, which is a special case of the above situation.

In Section 2 of this paper, the misclassification model introduced by Bross is extended to multinomial data, and the effects of misclassification on the sample estimates of the multinomial proportions are discussed. In Sections 3 and 4 maximum likelihood estimates of the multinomial proportions are given along with expressions for the asymptotic variances of these estimates. In Sections 5 through 9 the problem of determining the sample sizes \( n \) and \( N \) is discussed. A numerical example is given to illustrate the calculations involved in estimating \( p \), and in determining the optimum sample sizes. Finally, in the Appendix, the expressions for the maximum likelihood estimates of \( p \), and the corresponding variances are derived.

2. Effects of Misclassification on Estimating from Multinomial Data

Suppose a sample of units is drawn from some population and each unit is placed into one of \( r \) categories, which are denoted by the integers \( 1, 2, \ldots, r \). For each sampling unit we define the random variables \( T \) and \( F \) in the following
A DOUBLE SAMPLING SCHEME

$T = i$ if the sampling unit truly belongs to category $i$ ($i = 1, 2, \ldots, r$)

$F = j$ if the sampling unit is classified by the fallible device as being in category $j$ ($j = 1, 2, \ldots, r$)

Here, $T$ refers to the true measurement and $F$ refers to the fallible measurement. The marginal distributions of $T$ and $F$ are:

$$p_i = \Pr \{T = i\} \quad (2.1)$$
$$\pi_j = \Pr \{F = j\} \quad (2.2)$$

where $i$ and $j$ refer to the category numbers and $\sum_{i=1}^{r} p_i = \sum_{i=1}^{r} \pi_i = 1$. Throughout this paper $j$ and $i$ will take the values $1, 2, \ldots, r$ unless otherwise indicated.

To describe misclassification we define $\theta_{ij}$ to be the probability that a unit, which belongs to the $i$th category, is classified in the $j$th category. Thus:

$$\theta_{ij} = \Pr \{F = j \mid T = i\}. \quad (2.3)$$

It follows that $\sum_{j=1}^{r} \theta_{ij} = 1$ and

$$\pi_i = \sum_{j=1}^{r} p_j \theta_{ij}. \quad (2.4)$$

If a fallible classifier is used to classify units into each of the $r$ categories, the proportion of units observed in category $j$ is the maximum likelihood estimate of $\pi_j$. The bias which results if this proportion is used to estimate $p_i$ is therefore $\pi_i - p_i$.

We can obtain more insight into the nature of this bias by considering the multinomial estimation problem, in which we are interested in estimating $r - 1$ proportions as $r - 1$ binomial estimation problems. Suppose we consider the $j$th category and we define $\theta_j$ to be the probability of misclassifying a unit belonging to category $j$ and $\Phi_j$ to be the probability of classifying a unit, which does not belong to category $j$, as being in category $j$. Thus:

$$\pi_i = p_i (1 - \theta_j) + \Phi_j \quad (2.5)$$

In the case of binomial data where $r = 2$, Bross [1] has shown that

$$\pi = p (1 - \theta) + \theta \Phi$$

where $\theta$ and $\Phi$ are the corresponding probabilities of misclassification. As a result, the multinomial misclassification model is a generalized case of the binomial model.

The resulting bias in the estimate of $p_i$ is:

$$\pi_i - p_i = q_i \Phi_i - p_i \theta_i$$

and the relative bias is:

$$\text{Bias}/p_i = q_i \Phi_i / p_i - \theta_i. \quad (2.6)$$

It is easy to see that this bias can be serious for large $\Phi_i$ and small $p_i$. However,
it is possible to have a serious bias even when $\theta_i$ and $\Phi_i$ appear to be small. For instance, suppose $p_1 = .1$, $\theta_1 = .01$, and $\Phi_1 = .05$. Then, from (2.5), $r_1 = .114$ and the relative bias is 44%.

3. Estimation by Double Sampling

The situation which is hypothesized is as follows: it is possible but very costly to obtain the true value for each unit in the sample, but if a fallible classifier is used, biased estimates of the multinomial proportions $p_i$ will result. A compromise between these two procedures is a double sampling scheme which can be described as follows:

a) A sample of $n$ units is drawn and the true and fallible classifications, denoted by $T$, and $F_s$ ($s = 1, 2, \cdots, n$) respectively, are obtained for each unit. The following $r \times r$ contingency table summarizes the results:

\[
\begin{array}{cccc}
1 & 2 & \cdots & r \\
\begin{array}{cccc}
1 & n_{11} & n_{12} & \cdots & n_{1r} \\
2 & n_{21} & n_{22} & \cdots & n_{2r} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
r & n_{r1} & n_{r2} & \cdots & n_{rr} \\
\end{array}
\end{array}
\]

Here $n_{ij}$ equals the number of units in the sample whose true category is "i" and whose fallible category is "j". Note that:

\[
\sum_{i=1}^{r} n_{ij} = n_i, \\
\sum_{j=1}^{r} n_{ij} = n_j, \\
\sum_{j=1}^{r} \sum_{i=1}^{r} n_{ij} = n
\]

(3.2) (3.3)

b) A further sample of $N - n$ units is drawn and the fallible classifications $F_s$ ($s = n + 1, n + 2, \cdots, N$) is obtained for each unit. Let the vector of frequencies be

\[X = (X_1, X_2, \cdots, X_r)\]

where $X_j$ equals the number of units whose fallible category is "j". It follows that

\[\sum_{j=1}^{r} X_j = N - n\]

(3.4)

The maximum likelihood estimates of the multinomial proportions $p_i$, and the
misclassification probabilities $\theta_{ij}$ are derived in appendix A. The resulting estimates are:

$$\hat{\theta}_{ij} = \sum_{j=1}^{M} n_{ij} (X_i + n_{.i}) / (n_{.i}N)$$  \hspace{1cm} (3.5)$$

$$\hat{\theta}_{i} = (X_i + n_{.i}) n_{i.} / (n_{i.}NP_i)$$  \hspace{1cm} (3.6)$$

From (3.5) it is possible to give an intuitive justification for the estimate of $\hat{\theta}_{i}$:

The ratio $(X_i + n_{.i}) / N$ estimates the proportion of the $N$ units in the sample which have been placed in category $j$ by the fallible classifier. Of these units which have been placed in category $j$, a certain proportion truly belong to category $i$. An estimate of this proportion is given by $n_{ij} / n_{i.}$. Thus an estimate of the proportion of the $N$ units which truly belongs to category $i$ can be obtained by multiplying the proportion $(X_i + n_{.i}) / N$ by $n_{ij} / n_{i.}$ and summing over $j$.

It is possible to observe $n_{i.}$ to be zero from some value of $j$. This would imply that in the initial sample of $n$ units, no units were classified by the fallible device as being in category $j$. If this occurs, we define $n_{ij} / n_{i.} = 0$. This convention is consistent because, unless $n$ is extremely small, the event of observing $n_{i.} = 0$ will be rare except when $\pi_i$ is small. If the latter situation is true, then neglecting the term

$$n_{i.} (X_i + n_{.i}) / (n_{i.} N)$$

from (3.5) will have a negligible effect on this estimate of $\pi_i$.

Since $\hat{\theta}_{i}$ is the maximum likelihood estimate of $\pi_i$, it is asymptotically efficient. An expression for the asymptotic variance of $\hat{\theta}_{i}$, is derived in appendix B. The resulting expression is:

$$V_{i} = V(\hat{\theta}_{i}) = p_i q_i (1 - K_i) / n + p_i q_i K_i / N$$  \hspace{1cm} (3.7)$$

where

$$K_i = p_i \left( \sum_{j=1}^{M} \frac{\theta_{ij}}{\pi_i} - 1 \right) / q_i$$  \hspace{1cm} (3.8)$$

It is easy to show that $0 \leq K_i \leq 1$. Thus the variance of $\hat{\pi}_i$ ranges from $p_i q_i / n$ to $p_i q_i / N$. When there is no misclassification, $\theta_{ij} = 0$ for all $i \neq j$, $\theta_{ii} = 1$ and $\pi_i = p_i$. Thus, from (3.8), $K_i = 1$ and the variance is $p_i q_i / N$ as is to be expected. As $K_i$ increases, the asymptotic variance of $\hat{\pi}_{i}$ is given by (3.7) decreases. We define $K_i$ to be the coefficient of reliability corresponding to the $i$th category because it indicates how efficient the fallible measuring device is in estimating $\pi_i$.

The expression for the variance of $\hat{\theta}_{i}$, given by (3.7) is of the same form for the expression for the variance of $\hat{\pi}$ derived by Tenenbein [3] for the estimate of the binomial proportion $p$ by double sampling. The expression for the asymptotic variance of $\hat{\theta}$ was derived to be:

$$V(\hat{\theta}) = pq (1 - K) / n + pq K / N$$  \hspace{1cm} (3.9)$$

where $K = \text{coefficient of reliability}$. 
4. An Algorithm for Calculating the Estimates

A practical method for computing the values of $\hat{p}_i$ and $\hat{\theta}_{i,i}$ according to (3.5) and (3.6) from the data of table (3.1) is as follows:

a) Form the $r \times r$ matrix $A$ whose general term is $a_{ii} = n_{ii}/n_{i}$. This involves dividing every term in table (3.1) by its corresponding column sum.

b) Form the $r \times 1$ column vector $\pi$ whose $j$th entry is $\pi_j = (X_j + n_{i,j})/N$.

c) To obtain the $\hat{p}_i$ values form the product $P = A \times \pi$. $P$ is an $r \times 1$ column vector whose $i$th entry is $P_i$.

d) The values of $\hat{\theta}_{i,i}$ can be computed by the formula $\hat{\theta}_{i,i} = a_{i,i}/\pi_i$ where $a_{i,i}$, $\pi_i$, and $\hat{P}_i$ are the corresponding entries of $A$, $\pi$, and $P$ respectively.

This algorithm will be illustrated in the section 9 of this paper.

5. The Cost Function

In order to develop criteria for selecting $n$ and $N$ we must have some idea as to the cost of measurement; this will certainly be a consideration for determining the sample sizes. In this paper, we assume that the total cost of measurement, defined as $C$, is a linear function of $n$ and $N$, that is:

$$C = c_1n + c_2N = c_1n(Rf + 1)/(Rf)$$

(5.1)

where:

- $c_1 = \text{cost of obtaining one \"true\" measurement}$,
- $c_2 = \text{cost of obtaining one \"fallible\" measurement}$,
- $R = c_1/c_2 = \text{relative cost}$,
- $f = n/N = \text{subsampling fraction}$.

In practice, two situations may arise. An investigator might have a certain budget $C_0$ for measurement costs and he might want to choose $n$ and $N$ to minimize the precision of estimation. Or he might want to obtain a given precision of estimation at minimum cost. These two problems, referred to as the fixed cost and fixed precision problems respectively, are discussed in sections 6 and 7. The asymptotic variance of $\hat{p}_i$, defined as $V_i$ by (3.7) and (3.8) will be used as a measure of precision with which $p_i$ is estimated. The smaller the variance the greater is the precision.

In the next two sections, we assume that the parameters $c_1$, $c_2$, and the reliability coefficients $K_i$ given by (3.8) are known. Since the first two parameters are constants associated with the cost model, it is reasonable to assume that they are known. However, since the reliability coefficients are functions of the unknown parameters $p_i$ and $\theta_{i,i}$, it is not reasonable to assume that they will be known in practice. The problem of unknown reliability is discussed in section 8.

6. The Fixed Cost Problem

In the fixed cost problem, the budget for measurement costs is fixed at $C_0$. This implies that:
Subject to this constraint, the investigator ideally, would like to minimize $V_i$, given by (3.7), for $i = 1, 2, \cdots, r$. However, each of these functions depends upon two controllable variables, namely $n$ and $N$, subject to a cost constraint. Equivalently then, $V_i$ is a function of one controllable variable and thus this problem cannot be solved in general.

However, if the investigator wishes to minimize the precision of the estimate of $p_i$ in particular, the problem reduces to the following minimization problem:

Minimize: $V_i = p_i q_i (1 - K_i)/n + p_i q_i K_i/N$ for a given value of $i$. 

Subject to: $c_i n + c_i N \leq C_0$.

This problem was solved by Tenenbein [3] in connection with the binomial estimation problem. Using these results, the optimum value of $n$ and $N$ are given by:

$$n = n_s f_s/(f_s + 1)$$

$$N = n_s R/(R_i + 1)$$

where

$$f_i = \min (\sqrt{(1 - K_i)/(K_i R_i)}, 1)$$

$$n_s = C_0/c_1$$

$$R = c_1/c_2$$.

From (6.1) it is evident that $n_s$ represents the number of true measurements which can be used if no fallible measurements were used. The optimum value of $n$ is expressed as a fraction of $n_s$.

7. The Fixed Precision Problem

In the fixed precision problem, the investigator fixes the level of precision which he wishes the resulting estimates of $p_i$ to attain. He then chooses the values of $n$ and $N$ which minimize:

$$C = c_i n + c_i N$$

subject to the above precision constraint.

The required level of precision will vary from experimental situation to experimental situation. It will depend, among other things, upon the uses to which the estimates will be put. In some situations, the investigator may have difficulty in deciding upon the required level of precision. However, in many situations, he may be able to think of this precision in terms of the precision which can be attained from simple multinomial sampling of only true measurements. In this latter situation, he may state that the level of precision in the
double sampling scheme (n true and N fallible measurements) should be at least as great as the precision attained by using a sample of n true measurements.

If we again use the asymptotic variance of $\hat{p}_i$ as a measure of precision, the above problem can be stated as follows:

\[
\text{Minimize } C = c_1n + c_2N \tag{7.1}
\]

Subject to $V_i = p_{qi}(1 - K_i)/n + p_{qi}K_i/N \leq p_{qi}/n_i$.

The above constraints reduce to:

\[
\frac{(1 - K_i)}{n} + K_i/N \leq \frac{1}{n_i} \tag{7.2}
\]

for $i = 1, 2, \cdots, r$

Thus we wish to minimize a linear function of two variables, n and N, subject to r constraints. These r constraints can be reduced to one constraint by noting that if $K = \min (K_1, K_2, \cdots, K_r)$ then:

\[
\frac{(1 - K_i)}{n} + K_i/N \leq \frac{(1 - K)/n + K/N}{n_i} \tag{7.7}
\]

because $n < N$.

As a result the minimization problem given by (7.1) reduces to:

\[
\text{Minimize } C = c_1n + c_2N
\]

Subject to: $(1 - K)/n + K/N \leq 1/n_i$.

This problem was considered by Tenenbein [3] in connection with the binomial estimation problem. Using these results, the optimum values of n and N are given by:

\[
n = n_i(1 - K + Kf) \tag{7.3}
\]

\[
N = nn_iK/(n - n_i(1 - K))
\]

where

\[
f = \min \left( \sqrt{n - K}/KR, 1 \right)
\]

\[
R = c_i/c_2
\]

\[
K = \min (K_1, K_2, \cdots, K_r)
\]

The optimum value of n is expressed as a fraction of $n_i$.

Note that if we replace K by $K_1$, equation set (7.3) gives us the optimum value of n and N which minimize the measurement cost subject to the constraint that:

\[
V(\hat{p}_i) = p_{qi}(1 - K_i)/n + p_{qi}K_i/N \leq p_{qi}/n_i
\]

that is, subject to the constraint that the precision in estimating $p_i$ by double sampling is the same as the precision attained in estimating $p_i$ by simple multinomial sampling.
As indicated previously, the optimum formulae given by equations (6.2) and (7.3) depend upon the coefficient of reliabilities $K_i$, which are defined by (3.8). Since $K_i$ depends upon the parameters $p_i$, $\theta_i$, and $\pi_i$, it is not reasonable to assume that these parameters will be known unless prior experience is available. This creates a problem in applying these optimum formulae.

One possible solution to this problem is to take an initial pilot sample of $m$ units and classify each of these units by the true and the fallible measuring devices. This provides estimates of $p_i$, $\theta_i$, and $\pi_i$, and thus $K_i$ can be estimated from (3.8). Then estimates of the optimum value of $n$ and $N$ can be determined from the estimated reliabilities. This procedure is illustrated in the next section.

The problem of using this two stage scheme for determining the optimum values of $n$ and $N$ is the choice of the size of the pilot sample $m$. One can want $m$ to be sufficiently large so as to obtain reasonable estimates of the reliabilities $K_i$. On the other hand, if $m$ were too large, we run the risk of overshooting the true optimum value of $n$.

In connection with the binomial estimation problem, Tenenbein considered this problem in chapters 5 and 6 of [4]. This problem will be the subject of a future report.

In this section, a numerical example is given in order to illustrate the estimation procedure of section 4 and the sample size determination procedure discussed in section 8.

Suppose a firm wishes to examine the quality of units which it produces. The units produced are classified into three categories: defective, satisfactory, and superior. There are objective criteria for determining the category to which a unit belongs, but the expense of determining the exact category per sampling unit is $9, which is deemed high. A more inexpensive classification procedure would be subjective evaluation of the quality of the produced units by an inspector which costs $1 per sampling unit. However, this measurement technique is less precise because the inspection may tend to misclassify units.

Let $p_1$, $p_2$, and $p_3$ respectively denote the proportion of defective, satisfactory, and superior units produced by the company. In estimating these parameters by the double sampling presented in this paper, the firm wishes the precision to be at least as great as could be obtained from a multinomial sample of 100 true measurements.

This example corresponds to the fixed precision problem of section 7 where the cost function is $C = 9n + N$, the relative cost $R = 9$ and the precision level $n_0 = 100$. In order to determine $n$ and $N$ to minimize the cost subject to the precision constraint an initial sample of $m = 30$ units is taken and the true and fallible classifications are obtained for each unit. This will allow us to estimate the unknown reliabilities given by (3.8). We can then use the estimated reliabilities to estimate $n$ and $N$ by (7.3).

The data obtained from these 30 sampling units is given by the following $3 \times 3$ contingency table:
In this example, the inspector does not misclassify defective or superior units, but tends to misclassify satisfactory units. This effect may occur in practice because the categories “defective” and “superior” are extreme categories and the inspector may have an easier time of correctly classifying units in these two categories and a more difficult time of correctly classifying units which truly belong to the intermediate category labelled “satisfactory.”

We can estimate $P_i$ by the proportion of the 30 units which truly belong to category “$i$”. Thus:

$$P_i = \frac{\sum_i n_i}{30}$$

(9.2)

$\hat{P}_i$ can be estimated by the proportion of the 30 units which have been classified by the fallible device as being in category “$j$”. Thus:

$$\hat{P}_i = \frac{\sum_i n_{ij}}{30}$$

(9.3)

The parameter $\theta_{ij}$ can be estimated by taking the proportion of those units belonging to category $i$ which have been classified as being in category $j$. Thus:

$$\hat{\theta}_{ij} = \frac{\sum_i n_{ij}}{\sum_i n_i}$$

(9.4)

Substituting (9.2), (9.3), and (9.4) into (3.5) yields the following estimates of $K_i$:

$$K_1 = 0.583, \quad K_2 = 0.613, \quad K_3 = 0.573$$

and $K = \min \{K_1, K_2, K_3\} = 0.583$.

In order to obtain estimates of $n$ and $N$, we substitute $n_o = 100$, $R = 9$, and $K = 0.583$ into equation (7.3). The results of these computations yield $n = 58$ and $N = 207$. Now since an initial sample of $n = 30$ true-fallible measurements has already been taken, we need to take an additional sample of $n - m = 28$ units and obtain the corresponding true and fallible classifications. Since the optimum value of $N = 207$, and additional sample of $N - n = 149$ items is required for which only the fallible classifications are obtained. If in this example, $n$ were less than 50, then the optimum value of $n$ is exceeded. In this case, we would set $n = m$ and calculate $N$ from equation (7.3). Then an additional
sample of \( N - n = N - m \) items would be drawn and the corresponding fallible classifications would be obtained.

After taking the additional samples and combining these results with the results of the original 30 sampling units given by figure 9.1, the following data are observed:

<table>
<thead>
<tr>
<th>True Device</th>
<th>Defective (1)</th>
<th>Satisfactory (2)</th>
<th>Superior (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Defective</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Satisfactory</td>
<td>6</td>
<td>20</td>
<td>1</td>
</tr>
<tr>
<td>Superior</td>
<td>0</td>
<td>0</td>
<td>19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>First Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>N = 18</td>
</tr>
<tr>
<td>n = 20</td>
</tr>
<tr>
<td>N - n = 58</td>
</tr>
</tbody>
</table>

Using the algorithm described in section 4 of this paper, the values of \( \hat{p}_i \) are obtained by forming the matrix product \( A \times \pi \) where as before \( A \) is the matrix of elements \( a_{ii} = n_{ii}/n \), and \( \pi \) is the \( r \times 1 \) column vector of the values \( \pi_i = (X_i + n_{..i})/N \). The values of \( \hat{\theta}_{ii} \) are then obtained by the formula \( \hat{\theta}_{ii} = \pi_i a_{ii}/p_i \). The results of these calculations are tabulated in table A. Note that if the true classifier were not available the absolute value of the biases introduced by using \( \hat{\pi}_i \) to estimate the parameters \( p_i \) are .105, .121, and .016 respectively for categories 1, 2, and 3.

Using these estimates the estimated coefficients of reliability and the approximate standard errors of the \( \hat{\theta}_i \) can be computed from equation (3.7) and (3.8). The results are tabulated in table B below.

### Table A
**Estimated Values of \( \pi_i, p_i \) and \( \theta_{ii} \)**

<table>
<thead>
<tr>
<th>Category (i)</th>
<th>( \pi_i )</th>
<th>( p_i )</th>
<th>( \theta_{11} )</th>
<th>( \theta_{12} )</th>
<th>( \theta_{13} )</th>
<th>( \theta_{14} )</th>
<th>( \theta_{15} )</th>
<th>( \theta_{16} )</th>
<th>( \theta_{17} )</th>
<th>( \theta_{18} )</th>
<th>( \theta_{19} )</th>
<th>( \theta_{110} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.314</td>
<td>.209</td>
<td>.100</td>
<td>.00</td>
<td>0</td>
<td>.034</td>
<td>1</td>
<td>.221</td>
<td>.745</td>
<td>.034</td>
<td>.034</td>
<td>.034</td>
</tr>
<tr>
<td>2</td>
<td>.353</td>
<td>.474</td>
<td>.221</td>
<td>.745</td>
<td>.034</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>.333</td>
<td>.317</td>
<td>1</td>
<td>0</td>
<td>.034</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

### Table B
**Estimated Reliabilities and Standard Errors**

<table>
<thead>
<tr>
<th>Category (i)</th>
<th>Estimated Reliability (( \hat{\theta}_i ))</th>
<th>Approximate Standard Error of (( \hat{\theta}_i ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.577</td>
<td>.0408</td>
</tr>
<tr>
<td>2</td>
<td>.659</td>
<td>.0475</td>
</tr>
<tr>
<td>3</td>
<td>.944</td>
<td>.0110</td>
</tr>
</tbody>
</table>
10. Comparison Between Multinomial and Double Sampling

An alternate sampling procedure to the double scheme is simple multinomial sampling in which only the true classifications are obtained. In this section, these two sampling schemes are compared in order to provide information on the usefulness of the double sampling scheme.

With reference to this comparison, we define two reduction factors to correspond to the fixed cost and fixed precision problems respectively. Suppose in the fixed cost problem the investigator wishes to minimize the variance of $\hat{p}_t$ (for some fixed $t$) subject to the cost constraint. If we let $V_D$ be the minimum variance resulting from double sampling and $V_M$ be the minimum variance resulting from multinomial sampling, then:

$$\lambda_1 = 1 - \frac{V_D}{V_M}$$ (10.1)

Suppose, in the fixed precision problem the investigator wishes to minimize the measurement cost subject to the constraint that the variance of $\hat{p}_t$ is constrained by some value. If $C_D$ is the minimum cost of double sampling and $C_M$ is the minimum cost of multinomial sampling subject to the variance constraint, then

$$\lambda_2 = 1 - \frac{C_D}{C_M}$$ (10.2)

Tenenbein [3] derived expressions for $\lambda_1$ and $\lambda_2$ in the binomial estimation problem where the only parameter to be estimated is $p$. The results are that $\lambda_1 = \lambda_2$ and the common value $\lambda$ is:

$$\lambda = \frac{(1 - K + Kf)(R + W/R)}{R}$$

where

$$f = \min \left(\sqrt{(1 - K)/KR}, 1\right)$$ (10.3)

and

$$R = \text{coefficient of reliability}$$

Since the expression for the variance of the estimate of $\hat{p}_t$ is the same as the expression of the variance of $p$ with $p$ replaced by $\hat{p}_t$ and $K$ replaced by $K_t$ (compare 3.7 and 3.9), the above reduction factor $\lambda$ applies here with $K$ replaced by $K_t$. Thus $100\lambda$ can be interpreted as the percentage reduction in the variance resulting from double sampling or the percentage saving in cost resulting from double sampling.

Table 1 below shows values of $100\lambda$ as calculated by (10.3) for various values of the relative cost and the coefficient of reliability. Negative values indicate that multinomial sampling should be preferred, as is to be expected. The values of $\lambda$ increase as the reliability and the relative cost increase. When both the relative cost and the reliability coefficient tend to be high, the gains of the double sampling scheme are quite substantial. On the other hand, for some values of the relative cost and the reliability, the double sampling scheme seems hardly worthwhile (for example $R = 2$).
A DOUBLE SAMPLING SCHEME

TABLE 1
Percentage Reduction in Cost of Sampling and in Variance of Estimate of the Double Sampling Scheme over Multinomial Sampling (100X)
(Relative Cost (R))

<table>
<thead>
<tr>
<th>Reliability (K)</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>16</th>
<th>125</th>
<th>50</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-50.0</td>
<td>-20.0</td>
<td>-10.0</td>
<td>-6.2</td>
<td>-4.0</td>
<td>-2.0</td>
<td>-1.0</td>
</tr>
<tr>
<td>.05</td>
<td>-50.0</td>
<td>-20.0</td>
<td>-10.0</td>
<td>-6.2</td>
<td>-3.9</td>
<td>-1.3</td>
<td>0.6</td>
</tr>
<tr>
<td>.10</td>
<td>-55.0</td>
<td>-20.0</td>
<td>-10.0</td>
<td>-5.6</td>
<td>-2.4</td>
<td>1.3</td>
<td>3.9</td>
</tr>
<tr>
<td>.20</td>
<td>-50.0</td>
<td>-19.8</td>
<td>-7.3</td>
<td>-1.2</td>
<td>3.2</td>
<td>8.3</td>
<td>11.8</td>
</tr>
<tr>
<td>.30</td>
<td>-50.0</td>
<td>-17.0</td>
<td>-2.0</td>
<td>5.2</td>
<td>10.5</td>
<td>15.4</td>
<td>20.5</td>
</tr>
<tr>
<td>.40</td>
<td>-49.3</td>
<td>-15.8</td>
<td>0.0</td>
<td>13.0</td>
<td>18.8</td>
<td>25.3</td>
<td>29.8</td>
</tr>
<tr>
<td>.50</td>
<td>-45.7</td>
<td>-4.7</td>
<td>13.4</td>
<td>21.9</td>
<td>28.0</td>
<td>34.9</td>
<td>39.5</td>
</tr>
<tr>
<td>.60</td>
<td>-39.3</td>
<td>4.2</td>
<td>23.0</td>
<td>31.8</td>
<td>38.0</td>
<td>44.9</td>
<td>49.6</td>
</tr>
<tr>
<td>.70</td>
<td>-29.8</td>
<td>15.0</td>
<td>34.0</td>
<td>42.7</td>
<td>48.9</td>
<td>55.6</td>
<td>60.1</td>
</tr>
<tr>
<td>.80</td>
<td>-16.6</td>
<td>28.2</td>
<td>46.7</td>
<td>55.0</td>
<td>60.8</td>
<td>67.1</td>
<td>71.2</td>
</tr>
<tr>
<td>.90</td>
<td>2.6</td>
<td>45.2</td>
<td>62.0</td>
<td>69.4</td>
<td>74.4</td>
<td>79.7</td>
<td>83.1</td>
</tr>
<tr>
<td>.95</td>
<td>17.7</td>
<td>56.5</td>
<td>71.7</td>
<td>78.2</td>
<td>82.5</td>
<td>86.9</td>
<td>89.7</td>
</tr>
<tr>
<td>1.00</td>
<td>50.0</td>
<td>80.0</td>
<td>90.0</td>
<td>93.8</td>
<td>96.0</td>
<td>98.0</td>
<td>99.0</td>
</tr>
</tbody>
</table>

Note: A negative sign means that there is no gain in using the double sampling scheme over multinomial sampling; that is, multinomial sampling is preferred.

11. APPENDIX

A. Derivation of the Maximum Likelihood Estimates

In this section, we show that the maximum likelihood estimates of \( p_i \) and \( \theta_{ij} \) are given by

\[
\hat{p}_i = \frac{1}{N} \sum_{i=1}^{r} (X_i + n_{ii})n_{ii}/(N n_{ii})
\]

\[
\hat{\theta}_{ij} = (X_i + n_{ij})n_{ij}/(N n_{ij})
\]

At the first stage of the double sampling scheme, we observe \( n_{ii} \), the number of units whose true category is \( 'i' \) and whose fallible category is \( 'j' \). Consider \( n_{ij} \) for fixed \( i \). We can consider the \( r \times r \) contingency table, given by (3.1) as a \( 2 \times r \) contingency table, depending upon whether the true category of a unit is \( 'i' \) or not \( 'i' \).

Thus:
The joint distribution of the \( n_{ii} \) and \( n_{ij} - n_{ii} \) for \( j = 1 \) to \( r \) is multinomial with sample size \( n \). The corresponding cell probabilities of the \( n_{ii} \) are from (2.1) and (2.3):

\[
\Pr [T = i, F = j] = p_{i, \theta_{ij}} \quad (A1)
\]

The corresponding cell probabilities of the \( n_{ij} - n_{ii} \) are from (2.2) and (A1):

\[
\Pr [T \neq i, F = j] = \Pr [F = j] - \Pr [T = i, F = j] = \pi_r - p_{i, \theta_{ij}} \quad (A2)
\]

At the second stage we observe \( X_i \), the number of units out of \( N - n \) whose fallible category is "\( j \)". Thus, the joint distribution of the \( X_i \) is multinomial with sample size \( N - n \) and corresponding cell probabilities:

\[
\Pr [F = j] = \pi_r
\]

Since the second sample is independent of the first sample, the joint likelihood of \( n_{ii}, n_{ij} - n_{ii}, \) and \( X_i (j = 1, 2, \ldots, r, i \text{ fixed}) \) is proportional to

\[
L = \prod_{i=1}^{r} (p_{i, \theta_{ij}})^{n_{ii}} (\pi_r)^{n_{ij} - n_{ii}} \prod_{j=1}^{r} \pi_r^{X_i} \quad (A3)
\]

Let

\[
\lambda_{ij} = p_{i, \theta_{ij}} / \pi_r \quad (A4)
\]

Substituting (A4) into (A3), simplifying and taking logs of both sides yields:

\[
F = \log L - \sum_{i=1}^{r} n_{ii} \ln \lambda_{ii} + \sum_{i=1}^{r} (n_{ij} - n_{ii}) \ln (1 - \lambda_{ii}) + \sum_{i=1}^{r} (X_i + n_{i}) \ln n_r + \sum_{i=1}^{r} (X_i + n_{i}) \ln [1 - \sum_{i=1}^{r} \pi_r] \quad (A5)
\]

We wish to maximize \( F \) in order to find the maximum likelihood estimates of \( \lambda_{ij} \) and \( \pi_r \). Notice that:

\[
\sum_{i=1}^{r} \pi_r = 1 \quad (A6)
\]

Thus we wish to maximize (A5) subject to (A6) or thus to maximize:

\[
F = \sum_{i=1}^{r} n_{ii} \ln \lambda_{ii} + \sum_{i=1}^{r} (n_{ij} - n_{ii}) \ln (1 - \lambda_{ii})
+ \sum_{i=1}^{r} (X_i + n_{i}) \ln \pi_r + (X_r + n_{r}) \ln [1 - \sum_{i=1}^{r} \pi_r] \quad (A7)
\]
Setting \( \frac{\partial F}{\partial \lambda_{ij}} = 0 \) for \( j = 1, 2, \ldots, r \) and setting \( \frac{\partial F}{\partial \pi_j} = 0 \) for \( j = 1, 2, \ldots, r - 1 \) and solving the resulting equations for \( \lambda_{ij} \) and \( \pi_j \) yields:

\[
\hat{\pi}_j = \frac{(X_i + n_{ij})}{N} \quad \text{(A8)}
\]

\[
\hat{\lambda}_{ij} = \frac{n_{ij}}{n_i} \quad \text{(A9)}
\]

From (A4) it follows that

\[
p_i = \sum_{j=1}^{r} \lambda_{ij} \pi_i \quad \text{(A10)}
\]

Using the invariance property of maximum likelihood estimates, the maximum likelihood estimates, of \( p_i \) and \( \theta_i \), can be obtained by substitution of (A8) and (A9) into (A10) and (A4).

B. Asymptotic Variance of \( \hat{\theta}_i \)

The asymptotic variance of \( \hat{\theta}_i \) is given by:

\[
V(\hat{\theta}_i) = \frac{p_i q_i (1 - K_i)}{n} + \frac{p_i q_i K_i}{N} \quad \text{(B1)}
\]

where:

\[
K_i = p_i \left( \sum_{j=1}^{r} \frac{\theta_j^2}{\pi_j} - 1 \right) / q_i \quad \text{(B2)}
\]

In this section we present a lemma which leads to this result.

**Lemma 1.** Let \( \lambda_{ij}, \hat{\lambda}_{ij}, \) and \( \hat{\pi}_j \) be defined by (A4), (A8), and (A9) respectively.

Then:

a) \( V(\hat{\pi}_j) = \pi_j (1 - \pi_j) / N \) \quad \text{(B3)}

b) \( \text{COV} (\hat{\pi}_i, \hat{\pi}_k) = -a_i n_k / N \) \quad \text{(B4)}

c) The following results hold asymptotically (accurate to terms of order \( 1/n \)):

i) \( V(\hat{\lambda}_{ij}) = \lambda_{ij} (1 - \lambda_{ij}) / n \pi_i \) \quad \text{(B5)}

ii) \( \text{COV} (\hat{\lambda}_{ij}, \hat{\lambda}_{ik}) = \text{COV} (\hat{\lambda}_{ij}, \hat{\lambda}_{ik}) = \text{COV} (\hat{\lambda}_{ij}, \hat{\lambda}_{ik}) = 0 \) \quad \text{(B6)}

for all \( i \) and all \( j \neq k. \)

**Proof of (a) and (b).** Since \( X_i + n_i \) represents the total number of units out of \( N \) whose fallible classification is \( j \), the joint distribution of \( X_i + n_i \) is multinominal with sample size \( N \) and cell probabilities \( \pi_j \). Because \( \hat{\pi}_j = (X_i + n_i)_j / (N n_i) \), (B3), and (B4) follow immediately.

**Proof of (c).** Consider \( i \) fixed. Then the number of independent parameters is \( 2r - 1 \), namely \( \lambda_{1i}, \lambda_{2i}, \cdots, \lambda_{ri}, \pi_1, \pi_2, \cdots, \pi_r \). If we label these parameters by \( \alpha \), for \( i = 1, 2, \cdots, 2r - 1 \) the \((2r - 1) \times (2r - 1)\) information matrix whose general term is given by \( E[-\delta^2 F / \partial \alpha \delta \alpha] \) can be evaluated from (A7). Carrying out these calculations and inverting the resulting matrix yields the variance-covariance matrix of \( \lambda_{ij} \) and \( \pi_k \) for \( j = 1, 2, \cdots, r \) and \( k = 1, 2, \cdots, r - 1 \), from which (B5) and (B6) follow.
In order to evaluate the asymptotic variance of $\hat{\phi}_i$, we use the delta method (\cite{2}, page 321, equation 6a.2.6) which involves expanding $\hat{\phi}_i = \sum_{j=1}^{k} \lambda_{ij} \hat{\phi}_j$ in a Taylor's series about the true value $\phi_i = \sum_{j=1}^{k} \lambda_{ij} \pi_j$. The result of this expansion is (ignoring the zero covariance terms of lemma 1):

$$V(\hat{\phi}_i) = \sum_{i=1}^{r} V(\lambda_{ii}) (\partial \phi_i / \partial \lambda_{ii})^2 + \sum_{i=1}^{r} V(\pi_i) (\partial \phi_i / \partial \pi_i)^2$$

$$+ \sum_{i=1}^{r} \sum_{j=1}^{r} \text{COV}(\pi_i, \pi_j) (\partial \phi_i / \partial \pi_i)(\partial \phi_i / \partial \pi_j). \tag{B10}$$

From (A10) it follows that $\partial \phi_i / \partial \lambda_{ii} = \pi_i$ and $\partial \phi_i / \partial \pi_i = \lambda_{ij} = p_{ij} / \pi_i$. Substituting these results and equations (B3), (B4), and (B5) into equation (B10) and simplifying will yield the expression for the asymptotic variance of $\hat{\phi}_i$ as given by equations (B1) and (B2).

References


