



MICROCOPY RESOLUTION TEST CHART NATIONAL BUREAU OF STANDARDS -- 1963 -- A

γquí

FIE

E

ESTIMATION OF POPULATION PROPORTION AND OF PROBABILITIES OF MISCLASSIFICATION FROM IMPERFECT BINOMIAL DATA BY COMPARISON WITH RESULTS OF A STANDARD METHOD OF KNOWN INPERFECTION

Norman L. Johnson University of North Carolina Chapel Hill, North Carolina and

Samuel Kotz University of Maryland College Park, Maryland

NO0014-84-KD30/

August 1985



This document has been approved for public release and sale; its distribution is unlimited.

85 29 8 069

# ESTIMATION OF POPULATION PROPORTION AND OF PROBABILITIES OF MISCLASSIFICATION FROM IMPERFECT BINOMIAL DATA BY COMPARISON WITH RESULTS OF A STANDARD METHOD OF KNOWN IMPERFECTION

Norman L. Johnson University of North Carolina Chapel Hill, North Carolina and

Samuel Kotz University of Maryland College Park, Maryland

<u>ABSTRACT</u>: Observations from inspection by a 'test' method and a standard method are combined to provide estimators of population proportion, and of probabilities of misclassification for the test method. Results of Hochberg and Tenenbein (1983) and of Albers and Veldman (1984) are extended to the case where the standard method is not perfect, but its misclassification probabilities have known values. Both moment and maximum likelihood estimators are considered and some asymptotic properties of the resulting estimators are compared. *Additional Kaynovic method of misnature finemal distribution Statistics different apprendiation (mathinatics)*, *bace* 

Key Words and Phrases: Binomial distribution; information matrix; inspection errors; maximum likelihood; method of moments; statistical differentials.



### 1. INTRODUCTION

Suppose we have a large population, containing an unknown proportion, P, of individuals possessing a certain characteristic, which we will call 'nonconformance'. In a random sample, of size n, from this population, the distribution of the number, X, say, of nonconforming individuals will be binomial with parameters n,P so that

 $\Pr[X=x] = {n \choose x} P^{X}(1-P)^{n-X} (x=0,1,...,n).$ 

We will represent this, symbolically, as

 $X \frown Bin(n,P)$ , where  $\frown$  denotes "is distributed as".

If the individuals in the sample of size n are examined by an imperfect measuring device, which detects actual nonconformance with probability p, and (incorrectly) 'detects' nonconformance, when the individual is really not nonconforming, with probability p', then the distribution of Z, the number of individuals declared to be nonconforming, as a result of this inspection, will be binomial with parameters n, Pp + (1-P)p'. It is clear that the only parameter that can be estimated from observations on values of Z in independent samples is Pp + (1-P)p'.

Various methods have been suggested for obtaining data from which estimates of P, p and p' can be derived (e.g. Albers and Veldman (1984), Johnson and Kotz (1985)). Tenenbein (1970) suggested additional inspection of part of the sample by a perfect measuring device (for which p=1 and p'=0) and utilizing the resultant data. This method has been extended by Hochberg and Tenenbein (1983) to allow for inspection of a <u>further</u> sample, of size  $n_S$ , say, by the perfect measuring device (S).

In this paper, we study problems arising in this latter situation if the 'established' measuring device S is not perfect, but has known values  $p_S$ ,  $p'_S$  for p,p' respectively. For convenience, we will denote the (unknown) values of p,p' for the measuring device under test (T) by  $p_T, p'_T$  respectively. We will also assume (when necessary) that  $p_S >> p'_S$ and  $p_T >> p'_T$ .

## 2. ANALYSIS I (Moment Estimation)

As a consequence of the inspections we have the following sets of observations:

(i)  $n_S$  using S alone, with  $Z_S$  judged nonconforming (NC),

(ii)  $n_T$  using T alone, with  $Z_T$  judged NC,

(iii) n using both S and T, with results shown below:

s	# NC	# not NC
# NC	<sup>Z</sup> 11	<sup>Z</sup> 10
# not NC	<sup>2</sup> 01	<sup>Z</sup> 00

(# denotes 'number of'.) Evidently,  $Z_{11} + Z_{10} + Z_{01} + Z_{00} = n$ .

Under the assumption of random sampling from a population of effectively infinite size, we have that:

$$Z_{S}, Z_{T} \text{ and } Z = \begin{pmatrix} Z_{11} & Z_{10} \\ \\ Z_{01} & Z_{00} \end{pmatrix}$$
 are mutually independent; (1.1)

$$Z_{S} \cap Bin(n_{S}, \theta_{S})$$
 with  $\theta_{S} = p_{S}P + p'_{S}(1-P)$ ; (1.2)

$$Z_T \cap Bin(n_T, \theta_T)$$
 with  $\theta_T = p_T P + p_T'(1-P)$ ; (1.3)

$$\sum_{r} \sim Multinomial \left( n; \begin{pmatrix} \phi & \theta_{S} \phi \\ \theta_{T} \phi & 1 - \theta_{S} \theta_{T} \phi \end{pmatrix} \right)$$
(1.4)

and

with  $\phi = p_{S}p_{T}P + p_{S}'p_{T}'(1-P)$ , where  $\neg$  denotes "is distributed as".

Recall that  $p_S$  and  $p'_S$  have known values, and P is the (unknown) proportion of NC individuals in the population.

Hence

$$(n_{S} + n)\tilde{\theta}_{S} = Z_{S} + Z_{10} + Z_{11} - Bin(n_{S} + n, \theta_{S})$$
 (2.1)

$$(n_{T} + n)\tilde{\theta}_{T} = Z_{T} + Z_{01} + Z_{11} - Bin(n_{T} + n, \theta_{T})$$
 (2.2)

$$n\tilde{\phi} = Z_{11} \frown Bin(n,\phi)$$
(2.3)

so that  $\tilde{\theta}_{S} = \tilde{\theta}_{T}$  and  $\tilde{\phi}$  (as defined in (2.1)-(2.3)) are unbiased estimators of  $\theta_{S}$ ,  $\theta_{T}$  and  $\phi$  respectively.

Hence 
$$\tilde{\mathbf{p}} = (\mathbf{p}_{\mathrm{S}} - \mathbf{p}_{\mathrm{S}}')^{-1} (\tilde{\boldsymbol{\theta}}_{\mathrm{S}} - \mathbf{p}_{\mathrm{S}}')$$
 (3.1)

is an unbiased estimator of P. Although the estimators

$$\tilde{\mathbf{p}}_{\mathrm{T}} = (\tilde{\boldsymbol{\theta}}_{\mathrm{S}} - \mathbf{p}_{\mathrm{S}}')^{-1} (\tilde{\boldsymbol{\phi}} - \mathbf{p}_{\mathrm{S}}' \ \tilde{\boldsymbol{\theta}}_{\mathrm{T}})$$
(3.2)

and

d  $\tilde{p}_{T}' = (p_{S} - \tilde{\theta}_{S})^{-1} (p_{S} \tilde{\theta}_{T} - \tilde{\phi})$  (3.3)

are not unbiased estimators of  $p_T$  and  $p_T'$  respectively, the biases should not be large if sample sizes are adequate (see the example later in this section).

The variance-covariance matrix of the random variables in (2.1)-(2.3) is

$$\operatorname{Var}\left((n_{S}^{+n})\tilde{\theta}_{S},(n_{T}^{+n})\tilde{\theta}_{T}^{*},n\tilde{\phi}\right) = \begin{pmatrix} (n_{S}^{+n}) S^{(1-\theta_{S})} & n(\phi-\theta_{S}\theta_{T}) & n\phi(1-\theta_{S}) \\ n(\phi-\theta_{S}\theta_{T}) & (n_{T}^{+n})\theta_{T}^{*}(1-\theta_{T}) & n\phi(1-\theta_{T}) \\ n\phi(1-\theta_{S}) & n\phi(1-\theta_{T}) & n\phi(1-\phi) \end{pmatrix}$$
(4)

Hence (cf. (3.1))

$$Var(\tilde{P}) = (n_{S}+n)^{-1}(p_{S}-p'_{S})^{-2}\theta_{S}(1-\theta_{S})$$
(5.1)

and, using the method of statistical differentials (see, e.g. Johnson and Kotz (1969, Chapter 1, Section 7.5)) we obtain, after some algebraic manipulation:

$$\operatorname{var}(\tilde{p}_{T}) \stackrel{*}{=} p_{T}^{2} p^{-2} (p_{S}^{-} p_{S}^{+})^{-2} \left[ \left\{ n^{-1} \phi (1-\phi) - 2(n_{T}^{+}n)^{-1} p_{S}^{+} \phi (1-\theta_{T}) + (n_{T}^{+}n)^{-1} p_{S}^{+} 2\theta_{T} (1-\theta_{T}) \right\} p_{T}^{-2} - 2(n_{S}^{+}n)^{-1} \left\{ \phi (1-\theta_{S}) - n(n_{T}^{+}n)^{-1} p_{S}^{+} (\phi - \theta_{S} \theta_{T}) \right\} p_{T}^{-1} + (n_{S}^{+}n)^{-1} \theta_{S} (1-\theta_{S}) \right]$$
(5.2)

Δ

An approximate expression for  $var(\widetilde{p_T})$  is obtained from (5.2) by replacing  $p_T$  by  $p_T'$  and P by (1-P), and interchanging  $p_S$  and  $p_S'$ .

An approximate formula for the bias of  $\boldsymbol{p}_{T}$  is

$$E[\tilde{p}_{T}] - p_{T} \stackrel{*}{=} p_{T} \left\{ \frac{\operatorname{var}(\tilde{\theta}_{S})}{\left(\theta_{S} - p_{S}'\right)^{2}} - \frac{\operatorname{cov}(\tilde{\theta}_{S}, \tilde{\phi} - p_{S}', \tilde{\theta}_{T})}{\left(\theta_{S} - p_{S}'\right)\left(\phi - p_{S}', \theta_{T}\right)} \right\}$$
(6)

which, after some reduction, gives a proportional bias (i.e.  $100(bias)/p_T$ %)

$$\frac{100\{n_{T}p'_{S}(1-\theta_{S}) + n(1-p'_{S})\theta_{S}\}(\phi-\theta_{S}\theta_{T})}{(n_{S}+n)(n_{T}+n)(\theta_{S}-p'_{S})^{2}(\phi-p'_{S}\theta_{T})} \%$$
(7)

From (1.1)-(1.4)

$$\theta_{S} - p'_{S} = (p_{S} - p'_{S})P; \quad \phi - p'_{S} \quad \theta_{T} = p_{T}(p_{S} - p'_{S})P$$

 $\phi - \theta_{S} \theta_{T} = P(1-P)(p_{S} - p_{S}')(p_{T} - p_{T}')$ , so

and

the approximate proportional bias (7) is

$$\frac{100\{n_{T}p_{S}^{\prime}(1-\theta_{S}) + n(1-p_{S}^{\prime})\theta_{S}\}(1-P)(p_{T}^{-}p_{T}^{\prime})}{(n_{S}^{+}n)(n_{T}^{+}n)P^{2}(p_{S}^{-}p_{S}^{\prime})^{2}p_{T}}$$
(7)

which is positive and (since  $p_T < p_T$ ) less than

$$\frac{100 \ G(1-P)}{(n_{\rm S}+n)P^2(p_{\rm S}-p_{\rm S}')^2} \ \%$$
(8)

5

where

$$G = \frac{n_T}{n_T + n} p'_S(1 - \theta_S) + \frac{n_T}{n_T + n} (1 - p'_S) \theta_S, \qquad (9)$$

which lies between  $p_{S}^{\prime}(1\!-\!\theta_{S})$  and  $(1\!-\!p_{S}^{\prime})\theta_{S}^{\prime}$  .

Example. Using as 'typical' values of the probabilities  $p_S$ ,  $p'_S$  and P the values 0.9, 0.1 and 0.1 respectively we find that

$$G = (n_T + n)^{-1} (0.082n_T + 0.162n)$$

(so that G lies between 0.082 and 0.162) and the approximate proportional bias of  $\tilde{p}_{T}$  is between 0 and 1406.25 G( $n_{S}$ +n)<sup>-1</sup>%. Note that the upper limit is less than 227.8 ( $n_{S}$ +n)<sup>-1</sup>%, so if  $n_{S}$ +n>100 the approximate proportional bias is less than 2.28%. The next section contains a numerical assessment of formula (5.2), without specifying values of  $p_{T}$  and  $p_{T}^{*}$ .

### 3. Some Numerical Approximations

Utilizing the reasonable assumption that  $p_S >> p'_S$ , and neglecting terms in  $p'_S$  and  ${p'_S}^2$  in the numerator of (5.2) we find

$$\operatorname{Var}(\tilde{p}_{T}) \stackrel{::}{=} p_{T}^{2} p^{-2} (p_{S}^{-} p_{S}^{*})^{-2} \left\{ \frac{\phi(1-\phi)}{np_{T}^{2}} - \frac{2\phi(1-\theta_{S})}{(n_{S}^{+}n)p_{T}} + \frac{\theta_{S}(1-\theta_{S})}{n_{S}^{+}n} \right\}$$
(10)

Taking  $p_S = 0.9$ ,  $p'_S = 0.1$  so that  $\theta_S = 0.8P + 0.1$  and  $\phi = 0.9 p_T P + 0.1 p'_T (1-P)$ , we obtain from (10)

$$\operatorname{Var}(\tilde{p}_{T}) \stackrel{:}{=} \frac{p_{T}^{2}}{0.64P^{2}} \left[ \frac{\{0.9p_{T}P + 0.1p_{T}'(1-P)\}\{1-0.9p_{T}P - 0.1p_{T}'(1-P)\}}{n p_{T}^{2}} + \frac{(0.8P+0.1)(0.9-0.8P) - \{1.8p_{T}P + 0.2p_{T}'(1-P)\}(0.9-0.8P)p_{T}^{-1}}{n_{S} + n} \right]$$

Now taking P = 0.1, we find

$$\operatorname{var}(\tilde{p}_{T}) \stackrel{::}{=} \frac{p_{T}^{2}}{0.0064} \left[ \frac{0.09(p_{T}^{+}p_{T}^{+})\{1-0.09(p_{T}^{+}p_{T}^{+})\}}{n p_{T}^{2}} - \frac{0.1476}{n_{S}^{+}n} \cdot \frac{p_{T}^{+}}{p_{T}^{-}} \right]$$
(11)

$$\frac{14.375}{n} \quad (p_{T}+p_{T}')\{1-0.09(p_{T}+p_{T}')\} - \frac{23.06}{n_{S}+n} \cdot \frac{p_{T}'}{p_{T}}$$
(12)

Since  $0.09(p_T + p_T') \{1 - 0.09(p_T + p_T')\} < \frac{1}{4}$  (because  $0.09(p_T + p_T') < 1$ ) the right hand side of (6) is less than

$$(0.0256n)^{-1} < 39.1 n^{-1}$$
.

In the next section we will compare the asymptotic variances and covariances of  $\tilde{\theta}_S, \tilde{\theta}_T$  and  $\tilde{P}$  with those for maximum likelihood estimators  $\hat{\theta}_S, \hat{\theta}_T$  and  $\hat{P}$  of  $\theta_S, \theta_T$  and  $\hat{P}$  respectively.

### 4. ANALYSIS II (Maximum Likelihood Estimators)

The likelihood function of  $Z_S$ ,  $Z_T$  and Z is

 $\begin{pmatrix} {}^{n}S \\ {}^{s}S \\ {}^{s}S \end{pmatrix} \begin{pmatrix} {}^{n}T \\ {}^{s}T \\ {}^{s}T \end{pmatrix} \begin{pmatrix} {}^{n}T \\ {}^{s}T_{10}, {}^{z}T_{01}, {}^{z}T_{00} \end{pmatrix} \theta {}^{z}S_{(1-\theta_{S})}{}^{n}S^{-2}S {}^{z}T_{10}{}^{(\theta_{S}-\phi)}{}^{z}T_{10}{}^{(\theta_{T}-\phi)}{}^{z}T_{01}{}^{(1-\theta_{S}-\theta_{T}+\phi)}{}^{z}T_{00}{}^{(\theta_{T}-\phi_{S})}{}^{z}T_{10}{}^{z}T_{10}{}^{(\theta_{T}-\phi_{S})}{}^{z}T_{10}{}^$ 

Equating derivatives of the log-likelihood to zero gives the following equations for  $\hat{\theta}_{S}$ ,  $\hat{\theta}_{T}$  and  $\hat{\phi}$ :

$$\frac{z_{\rm S}}{\hat{\theta}_{\rm S}} - \frac{n_{\rm S}^{-2} s_{\rm S}}{1 - \hat{\theta}_{\rm S}} + \frac{z_{10}}{\hat{\theta}_{\rm S}^{-\hat{\phi}}} - \frac{z_{00}}{1 - \hat{\theta}_{\rm S}^{-\hat{\theta}} r^{+\hat{\phi}}} = 0$$
(13.1)

$$\frac{Z_{\rm T}}{\hat{\theta}_{\rm T}} - \frac{n_{\rm T}^{-Z_{\rm T}}}{1 - \hat{\theta}_{\rm T}} + \frac{Z_{01}}{\hat{\theta}_{\rm T}^{-\hat{\phi}}} - \frac{Z_{00}}{1 - \hat{\theta}_{\rm S}^{-\hat{\theta}}_{\rm T}^{+\hat{\phi}}} = 0$$
(13.2)

$$\frac{z_{11}}{\hat{\phi}} - \frac{z_{10}}{\hat{\theta}_{S} - \hat{\phi}} - \frac{z_{01}}{\hat{\theta}_{T} - \hat{\phi}} + \frac{z_{00}}{1 - \hat{\theta}_{S} - \hat{\theta}_{T} + \hat{\phi}} = 0$$
(13.3)

The information matrix is

$$\underbrace{ \underbrace{ \underbrace{ \underbrace{ \underbrace{ \frac{n_{S}}{\theta_{S}^{(1-\theta_{S})}} + \frac{n(1-\theta_{T})}{(\theta_{S}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}}_{ \frac{n_{T}^{-1-\theta_{S}^{-\theta}T^{+\phi}}} + \frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}} - \frac{\frac{n(1-\theta_{T})}{(\theta_{S}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}}}{\frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}} - \frac{\frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}}}{\frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}} - \frac{\frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}}{\frac{n(\theta_{S}^{-\phi}T^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}{\theta_{T}^{-\phi}T^{-\phi}T^{-\phi}}} + \frac{\frac{n(1-\theta_{S})}{(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}T^{+\phi})}}}{\frac{n(\theta_{S}^{-\phi}T^{-\phi}$$

The determinant is

$$|\underline{v}^{-1}| = \frac{n}{\phi(\theta_{S}^{-\phi})(\theta_{T}^{-\phi})(1-\theta_{S}^{-\theta}-\theta_{T}^{+\phi})} \left[ \frac{n_{S}n_{T} \gamma}{\theta_{S}\theta_{T}(1-\theta_{S})(1-\theta_{T})} + n(n_{S}^{+}+n_{T}^{+}+n_{S}^{+}) \right]$$
  
with  $\gamma = \theta_{S}\theta_{T}(1-\theta_{S}^{-}-\theta_{T}^{+}+\phi) - \phi(\phi-\theta_{S}\theta_{T}).$  (15)

and from the asymptotic variance-covariance matrix V we obtain

$$\operatorname{Var}(\hat{\theta}_{S}) \stackrel{:}{=} \frac{n}{|\nabla^{-1}|} \left\{ \frac{n_{T}}{|\theta_{T}(1-\theta_{T})} \left[ \frac{1}{\phi} + \frac{1}{\theta_{S}-\phi} + \frac{1}{\theta_{T}-\phi} + \frac{1}{1-\theta_{S}-\theta_{T}+\phi} \right] + n \left[ \frac{1}{\theta_{T}-\phi} + \frac{1}{1-\theta_{S}-\theta_{T}+\phi} \right] \left\{ \frac{1}{\phi} + \frac{1}{\theta_{S}-\phi} \right\} \right\}$$

$$\stackrel{:}{=} \frac{\theta_{S}(1-\theta_{S})}{(\gamma n_{S}^{-1} + \delta N^{-1})/(\gamma + \delta)}$$

where  $N = n_S + n_T + n$  (= total number of observations) and

$$\delta = \frac{nN}{n_{\rm S}n_{\rm T}} \theta_{\rm S} \theta_{\rm T} (1-\theta_{\rm S}) (1-\theta_{\rm T})$$

The MLE of P is

$$\hat{P} = (P_{S} - P'_{S})^{-1} (\hat{\theta}_{S} - P'_{S})$$
(16)

The asymptotic efficiency of  $\tilde{P}$  (see (3.1)) is the same as that of  $\tilde{\theta}_{\text{S}}^{},$  which is

$$100(n_{S}^{+n})(\gamma n_{S}^{-1} + \delta N^{-1})/(\gamma + \delta)$$
(17)

7

(14)

Taking  $p_S = p_T = 0.9$ ,  $p'_S = p'_T = 0.1 = P$  and  $n_S = n_T = n$  (=  $\frac{1}{3}$  N) we find  $\gamma = 0.0184680$  and  $\delta = 0.0653573$ , so (17) becomes

$$100(n_{s}+n) (0.2203 n_{s}^{-1} + 0.7797 N^{-1})$$
$$= 2(0.2203 + 0.2599) = 96.04\%.$$

The asymptotic variance of the MLE  $\hat{\varphi}$  is

$$\operatorname{var}(\hat{\phi}) \stackrel{:}{=} \frac{1}{n} \frac{\delta \phi(1-\phi) - \delta \phi^2 N^{-1} \{ n_{S} \theta_{S}^{-1} (1-\theta_{S}) + n_{T} \theta_{T}^{-1} (1-\theta_{T}) \} + \phi(\theta_{S}-\phi) (\theta_{T}-\phi) (1-\theta_{S}^{-\theta} T^{+\phi})}{\delta + \gamma}$$

On the other hand, recalling that  $var(\tilde{\phi}) = n^{-1}\phi(1-\phi)$ , we find for the numerical values of the parameters used above, that the asymptotic efficiency of the moment estimator  $\phi$  is

$$100 \times \underbrace{0.0653573 \times 0.09\{0.91 - (2/3) \times 0.09 \times (0.18)^{-1} \times 0.82\} + 0.09 \times 0.09^{2} \times 0.73}_{0.09 \times 0.91 \ 0.0653573 + 0.0184680\}$$

$$= 100 \times \frac{0.0037449 + 0.0005322}{0.0068653} = 62.30\%$$

The markedly lower asymptotic efficiency of  $\tilde{\phi}$  is associated with the fact that it does not utilize the information on values of  $\theta_S$  and  $\theta_T$  which is available from the other  $(n_S + n_T)$  observations. Some support for this statement comes from the asymptotic efficiency of  $\tilde{\phi}$  if the values of  $\theta_S$  and  $\theta_T$  are <u>known</u>. This is

$$100 \times \frac{(\theta_{\rm S}^{-\phi})(\theta_{\rm T}^{-\phi})(1-\theta_{\rm S}^{-\theta}-\theta_{\rm T}^{+\phi})}{\gamma(1-\phi)}$$
(19)

With the numerical values of  $\theta_S$ ,  $\theta_T$  and  $\phi$  which we have been using above this would give an asymptotic efficiency of only <u>35.18</u>%.

8

(18)

### 5. CORRELATIONS

The asymptotic covariances between the MLE's  $\hat{\theta}_S$  and  $\hat{\theta}_T$ , and between  $\hat{\theta}_S$  and  $\hat{\phi}$ , are

$$\operatorname{cov}(\hat{\theta}_{S},\hat{\theta}_{T}) \stackrel{*}{=} N^{-1} \delta(\phi - \theta_{S} \theta_{T}) / (\gamma + \delta)$$
<sup>(20)</sup>

and

$$\operatorname{cov}(\hat{\theta}_{S}, \hat{\phi}) \neq N^{-1} \delta \phi \{ \operatorname{nm}_{T}^{-1} (1 - \phi \theta_{S}^{-1}) + 1 - \theta_{S} \} / (\gamma + \delta)$$

$$(21)$$

Since  $0 \leq \theta_S \theta_T \leq \phi \leq \theta_S \leq 1$ , the covariances cannot be negative. It is of interest to compare (20) with the covariance between the moment estimators  $\tilde{\theta}_S$  and  $\tilde{\theta}_T$ , which is

$$\operatorname{cov}(\tilde{\theta}_{S},\tilde{\theta}_{T}) = n(n_{S}+n)^{-1}(n_{T}+n^{-1}(\phi-\theta_{S}\theta_{T})$$
(22)

and (21) with the covariance between  $\tilde{\theta}_{S}$  and  $\tilde{\phi},$  which is

$$\operatorname{cov}(\tilde{\theta}_{S},\tilde{\phi}) = (n_{S}+n)^{-1}\phi(1-\theta_{S})$$
(23)

Using our illustrative parameter values ( $p_S=p_T=0.9$ ;  $p_S'=p_T'=0.1=P$ ;  $n=n_S=n_T$ ) we obtain the following values for correlations:

> $\operatorname{corr}(\tilde{\theta}_{S}, \tilde{\theta}_{T}) = 0.195; \quad \operatorname{corr}(\tilde{\theta}_{S}, \tilde{\phi}) = 0.475$  $\operatorname{corr}(\hat{\theta}_{S}, \hat{\theta}_{T}) \doteq 0.021; \quad \operatorname{corr}(\hat{\theta}_{S}, \hat{\phi}) \doteq 0.513$

Note that if n=0 (and  $n_S$ ,  $n_T > 0$ ) then

$$\operatorname{corr}(\hat{\theta}_{S},\hat{\theta}_{T}) = \operatorname{corr}(\hat{\theta}_{S},\hat{\theta}_{T}) = 0.$$

Also, if  $n_s = n_T = 0$  (and n > 0) then

$$\operatorname{corr}(\tilde{\theta}_{S}, \tilde{\theta}_{T}) = \operatorname{corr}(\hat{\theta}_{S}, \hat{\theta}_{T}) = (\phi - \theta_{S} \theta_{T}) \{\theta_{S} \theta_{T} (1 - \theta_{S}) (1 - \theta_{T})\}^{-\frac{1}{2}}$$

$$(= 0.390)$$

and

$$\operatorname{corr}(\tilde{\theta}_{S},\tilde{\phi}) = \operatorname{corr}(\hat{\theta}_{S},\hat{\phi}) = \left[\phi(1-\theta_{S})/\{\theta_{S}(1-\phi)\}\right]^{\frac{1}{2}}$$

$$(= 0.147)$$

(Numbers in parentheses are values corresponding to  $p_S = p_T = 0.09$ ,  $p'_S = p'_T = 0.1 = p$ .

### Acknowledgement

Dr. Samuel Kotz's work was supported by the U.S. Office of Naval Research under contract N00014-84-K-0301.

### REFERENCES

- Albers, W. and Veldman, H.J. (1984) Adaptive estimation of binomial probabilities under misclassification, <u>Statist. Neerland.</u>, 38, 233-247.
- Hochberg, Y. and Tenenbein, A. (1983) On triple sampling schemes for estimating from binomial data with misclassification errors, <u>Commun. Statist. - Theor.</u> <u>Math.</u>, 12, 1523-1533.
- Johnson, N.L. and Kotz, S. (1969) <u>Distributions in Statistics: Discrete</u> <u>Distributions</u>, Wiley, New York.
- Johnson, N.L. and Kotz, S. (1985) Some tests for detection of faulty inspection, Statist. Hefte, 26, 19-29.
- Tenenbein, A. (1970) A double sampling scheme for estimating from binomial data with misclassification, J. Amer. Statist. Assoc., 65, 1350-1362.

	I WINGTHOOM	(FILD)				Λ. Λ	109.51	1/
ECURITY	CLASSIF. CATI	ION OF THI	S PAGE			AD.A	15 1 12	Le <u> </u>
				REPORT DOCUME	NTATION PAGE	E		
Nonclassified			16. RESTRICTIVE MARKINGS					
28. SECURITY CLASSIFICATION AUTHORITY			3. DISTRIBUTION/A	VAILABILITY	FREPORT			
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE								
4. PERFOR	MING ORGAN	IZATION P	REPORT NUM	BER(S)	S. MONITORING OR	GANIZATION R	EPORT NUMBER	5)
UN	4/MSS/85	5/7					•	
SA NAME (	OF PERFORM	ING ORGA	NIZATION	Sb. OFFICE SYMBOL	7. NAME OF MONI	TORING ORGAN	ZATION	
Univ. of Maryland					U.S. Off	fice of Nav	al Research	
Se. ADORESS (City, State and ZIP Code)			<u> </u>	7b. ADDRESS (City,	State and ZIP Cod	le)		
Dept	c of Man	agemer	nt and S	Statistics				•
Col]	lege_Par	<u>rk</u> , MD	<u>20742</u>	•				
ORGAN	OF FUNDING	SPONSORI	NG	85. OFFICE SYMBOL	9. PROCUREMENT	NSTRUMENT IO	ENTIFICATION N	UMBER
	ONR				N 00014-8	34-K-301		
C. ADORE	SS (City, State	and ZIP Co	de)	····	10. SOURCE OF FUR	DING NOS.		
Stat. and Probability P Office of Naval Researc			ability H 1 Researc	rogram ch	PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT
Arlington, VA 22217			22217	,				
11. TITLE (Include Security Classification)"Estimation of pop			nation of popula	tion proportio	on and of p	robabilitie	s of mis- ard method	
<u>ciassii</u>	tication 1	ITOM IM	perfect t	monital data by	Comparison w	Lai resares	y or a scana	uru metarou
2. PERSON	NAL AUTHOR	(\$)	of known	imperfection"		· · · · · · · · · · · · · · · · · · ·		-
2. PERSON	NALAUTHOR	nd KOTZ	of known <u>Samuel</u>	imperfection"				
2. PERSOI JOHNSON 34. TYPE TECH	NAL AUTHOR <u>N.L.</u> OF REPORT INICAL	nd KOTZ	of known Samuel 135. TIME C	imperfection" overed 0/84 to 9/85	14. DATE OF REPOR August 19	ат (Yr., Mo., Dey, 85	15. PAGE C 11 pag	COUNT
2. PERSON JOHNSON ISA. TYPE TECH	NAL AUTHOR <u>N. N. L. aj</u> of report <u>INICAL</u> <u>EMENTARY</u> NO	NS) nd KOTZ	of known Samuel 136. TIME C FROM	imperfection" overed 0/84 to 9/85	14. DATE OF REPOR August 19	ат (Yr., Mo., Dey, 85	15. PAGE C 11 pag	COUNT CS
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE	NAL AUTHOR <u>N. N.L. ai</u> OF REPORT INICAL IMENTARY NO	NS) nd KOTZ	of known Samuel 135. TIME C FROM	imperfection" overed 9/84 to 9/85	14. DATE OF REPOR August 19	ат (Yr., <i>Mo., Dey</i> , 85	15. PAGE C 11 pag	OUNT CS
12. PERSOI JOHNSON IJL TYPE TECH 6. SUPPLE	NAL AUTHOR <u>N. N.L. a</u> OF REPORT INICAL MENTARY NO COSATI	CODES	of known <u>Samuel</u> 135. TIME C FROM	imperfection" overed )/84 to 9/85	14. DATE OF REPOR August 19	at (Yr., Mo., Dey, 85	ify by block number	
2. PERSOI JOHNSON 13. TYPE TECH 6. SUPPLE	NAL AUTHOR	CODES	of known <u>Samuel</u> 135. TIME C FROM <u></u>	imperfection" <u>overed</u> <u>0/84 to 9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu	14. DATE OF REPOR August 19 ontinue on reverse if no ibution; info m likelihood;	85 85 mation maintenant	ify by block number Trix; inspections; st	es tion atistical
2. PERSOI JOHNSON 34. TYPE 134. TYPE 16. SUPPLE 17. FIELD	NAL AUTHOR N. N.L. AI OF REPORT INICAL MENTARY NO COSATI GROUP	CODES	of known <u>Samuel</u> 135. TIME C FROM B. GR.	imperfection" overed 0/84 to 9/85 18 SUBJECT TERMS (C binomial distr errors; maximu differentials.	14. DATE OF REPOR August 19 ontinue on reverse if no ibution; info m likelihood;	t (Yr. Mo., Dey, 85 mation main mation main method of	<i>ty by block number</i> <i>Trix; inspec</i> moments; st	tion atistical
12. PERSOI JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRA	NAL AUTHOR N. N.L. AI OF REPORT INICAL MENTARY NO COSATI GROUP ACT 'Continue	CODES	of known <u>Samuel</u> 135. TIME C FROM 8. GR. 8. GR.	imperfection" overed 0/84 to 9/85 18. SUBJECT TERMS (C binomial distr errors; maximu differentials. 4. identify by block number	14. DATE OF REPOR August 19 ontinue on reverse if no ibution; info m likelihood;	eresson, and idea mation mat method of	<i>ty by block number</i> <i>trix; inspec</i> moments; st	tion atistical
12. PERSOI JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. ABSTRA Obset to prov for the (1984) misclas estimat	COSATI COSATI GROUP COSATI GROUP ACT (Continue ervations vide estin e test me are extensifications tors are	codes codes codes codes conside	of known <u>Samuel</u> 13b. ТIMÉ C FROM FROM B. GR. 8. GR. 6/ necessery end nspection of popula Results of the case abilities red and	imperfection" overed <u>9/84</u> to <u>9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; 19 ibution; 19 m likelihood; , thod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of	andard methabilities ( 83) and perference ( 83) and perference ( 83) and mathabilities ( 83) and perference ( 84) and mathabilities ( 85) and perference ( 85) and mathabilities ( 85) and perference ( 85) and mathabilities ( 85) and perference ( 85) and perferen	nod are comb of misclassi Albers and ect, but its aximum likel ting estimat	rtion atistical fication Veldman ihood ors are
12. PERSOI JOHNSON 13. TYPE TECH 16. SUPPLE 16. SUPPLE 17. FIELD 19. ABSTR 0bse to prov for the (1984) misclas estimat compare	NAL AUTHOR N. N.L. AI OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extensistications tors are ed.	codes oration codes from i mators thod. nded to on prob conside	of known <u>Samuel</u> 13b. TIME C FROM B. GR. 8. GR. 8. GR. 7. necessary and of popula Results of the case abilities red and	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; info m likelihood; , , thod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of	andard methabilities of s not performent and mainting the result	nod are comb of misclassi Albers and ect, but its aximum likel ting estimat	tion atistical fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRA Obset to prov for the (1984) misclass estimat compare	NAL AUTHOR N. N.L. AI OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extended ssification tors are ed.	codes codes codes codes conside	of known <u>Samuel</u> 13b. TIME C FROM B. GR. B. GR. Inspection of popula Results the case abilities red and	imperfection" overed 0/84 to 9/85 18 SUBJECT TERMS (C binomial distr errors; maximu differentials. didentify by block number h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; info m likelihood; , , , , , , , , , , , , , , , , , , ,	andard methabilities of 83) and of s not perfer the result	hod are combof misclassi Albers and ect, but its aximum likel ting estimat	fication vined fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRA Obset to prove for the (1984) misclass estimat Compare	NAL AUTHOR N. N.L. AJ OF REPORT NICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide esting are extensifications tors are ed.	codes oration codes from i mators thod. nded to on prob	of known Samuel 135. TIME C FROM B. GR. a. GR. a. GR. a. GR. a. GR. a. GR. c. C.	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; info m likelihood; , , , and of prob Tenenbein (19 dard method i ues. Both mo properties of	andard meth andard meth abilities of s not perfe- ment and ma the result	nod are comb of misclassi Albers and ect, but its aximum likel ting estimat	rion atistical fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTR (1984) Misclas estimat Compare	NAL AUTHOR N. N.L. AJ OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extended ssification tors are ed.	codes oration codes codes su on reverse i from i mators thod. nded to on prob conside	of known <u>Samuel</u> 13b. TIME C FROM B. GR. B. GR. Inspection of popula Results of the case abilities red and s	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; info m likelihood; , , thod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of	andard methabilities of 83) and of 83) and of 83) and of the result	<i>ty by block number</i> <i>Til pag</i> <i>trix; inspec</i> moments; st nod are comb of misclassi Albers and ect, but its aximum likel ting estimat	tion atistical fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. ASSTR 19. ASSTR 19	NAL AUTHOR N. N.L. A) OF REPORT NICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extensification tors are ed.	codes codes codes from i mators thod. nded to on prob	of known Samuel 135. TIME C FROM B. GR.	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; 19 ibution; 19 m likelihood; , thod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of	andard meth abilities of s not perfe- ment and ma the result	is page of line pa	rtion atistical fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRI Obse to prov for the (1984) misclass estimat compare	NAL AUTHOR N. N.L. AI OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extended ssification tors are ed.	CODES CODES CODES CODES CONTRUETS CONTRUETS CONTRUETS CONTRUETS CONTRUETS CONTRUETS CONTRUETS CONTRUETS CONTRUETS CODES	of known <u>Samuel</u> 13b. TIME C FROM B. GR. B. GR. If necessary and of popula Results of the case abilities red and s	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>1a SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 199 ibution; info m likelihood; , , , , , , , , , , , , , , , , , , ,	andard methabilities of andard methabilities of 83) and of s not performent and ma the result	Albers and ect, but its aximum likel	tion atistical fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRI Obset to prov for the (1984) misclas estimat compare	NAL AUTHOR N. N.L. AJ OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are exte ssifications tors are ed. BUTION/AVA FIED/UNLIMI	CODES CODES CODES CONTATION CODES CONTATION CODES CONTATION CODES CONTATION CODES CO	of known <u>Samuel</u> 13b. TIME C FROM B. GR. B. GR. C. C. C. C. C. C. C. C. C. C. C. C. C.	imperfection" overed <u>9/84 ro 9/85</u> <u>18 SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 ibution; info m likelihood; , , , , , , , , , , , , , , , , , , ,	andard meth andard meth abilities ( 83) and of s not perfe ment and ma the result	ty by block number in page frix; inspection moments; st nod are comb of misclassi Albers and ect, but its aximum likel ting estimat	rtion atistical fication Veldman ihood ors are
12. PERSOI JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. ABSTRI Obse to prov for the (1984) misclass estimat compare	NAL AUTHOR N.L. AI OF REPORT INICAL IMENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are extend ssifications tors are ed. BUTION/AVA FIED/UNLIMIT OF RESPONS	CODES CO	of known <u>Samuel</u> 13b. TIME C FROM FROM 8. GR. 8. GR. 7. necessary and nspection of popula Results of the case abilities red and s of ABSTRAC AME AS APT.	imperfection" OVERED 0/84 TO 9/85 18. SUBJECT TERMS (C binomial distr errors; maximu differentials. didentify by block number h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	<ul> <li>14. DATE OF REPOR August 199 August 199</li> <li>chod and a st ibution; info m likelihood;</li> <li>chod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of</li> <li>21. ABSTRACT SECU Nonclassif</li> <li>22b. TELEPHONE N</li> </ul>	andard meth abilities of andard meth abilities of 83) and of s not perfer ment and ma the result	Ty by block number Tix; inspective moments; st nod are combo of misclassi Albers and ect, but its aximum likel ting estimat CATION 22c. OFFICE SYM	tion atistical oined fication Veldman ihood ors are
12. PERSON JOHNSON 13. TYPE TECH 16. SUPPLE 17. FIELD 19. AGSTRI (1984) misclass estimat compare 10. DISTRIC INCLASSIN 12. NAME Samu	NAL AUTHOR N.L. AJ OF REPORT INICAL MENTARY NO COSATI GROUP ACT (Continue ervations vide estin e test me are exter ssification tors are ed. BUTION/AVAR FIED/UNLIMIT OF RESPONSI- 121 Kotz	CODES CODES CODES CODES CONTATION CODES CONTATION CODES CODES CONTATION CODES CODE	of known <u>Samuel</u> 13b. TIME C FROM B. GR.	imperfection" overed <u>9/84</u> <u>ro</u> <u>9/85</u> <u>1a SUBJECT TERMS (C</u> binomial distr errors; maximu differentials. <i>identify by block number</i> h by a 'test' me ation proportion of Hochberg and e where the stan s have known val some asymptotic	14. DATE OF REPOR August 19 August 19 ibution; Info m likelihood; thod and a st , and of prob Tenenbein (19 dard method i ues. Both mo properties of 21. AESTRACT SECU Nonclassif 22b. TELEPHONE N (Include Area Co (301) 454-6	andard meth andard meth abilities of 83) and of s not perfe- ment and ma the result ied	Albers and ect, but its aximum likel ting estimat	fication dined fication Veldman ihood ors are

# FILMED

EN

11-85

DTIC