ESTIMATION BY MOMENTS IN A MODEL OF FAULTY INSPECTION

Mao ShiSong  Samuel Kotz  Norman L. Johnson
East China Normal  University of  University of North
University, Shanghai  Maryland, College Park  Carolina, Chapel Hill

Key Words and Phrases: Attribute inspection; binomial distribution; factorial moments; hypergeometric distribution; mixtures; moment method.

ABSTRACT
Methodology developed by Blischke (Ann. Math. Statist. 33 (1962), 444-54) is applied to estimate the parameters in a model of faulty inspection, and to obtain approximate formulae for the variances of these estimators.

June 1985
1. INTRODUCTION

Recent papers (Johnson et al. (1980), Johnson & Kotz (1983), Kotz & Johnson (1982)) have developed distributions of observed numbers of apparently defective items when sample inspection is imperfect, resulting in some defectives not being observed as such, and possibly some non-defective being described as 'defective' ("false positives"). Although these results are of interest, some more practical problems arise when it is desired to test whether the inspection is faulty or to estimate the degree of imperfection. In Johnson and Kotz (1985) some tests for detection of faulty inspection were investigated. The present paper is devoted to the estimation aspects of the problem. We will consider here the simplest form of inspection by attributes, assuming lot size to be, effectively, infinite. Each individual in a random sample of size \( n \) is examined and a decision reached as to whether or not it is 'nonconforming' (NC). Ideally, of course, such decisions should be completely free of error, but, as is well-known, this is often not the case. As a model of faulty inspection, we introduce two parameters

\[ p = \Pr[\text{individual declared NC} \mid \text{individual is NC}] \]

\[ p' = \Pr[\text{individual declared NC} \mid \text{individual is not NC}] \]

and suppose we wish to estimate these parameters. The proportion, \( P \), of NC individuals in the lot is unknown, and plays the role, in this context, of a nuisance parameter.
It is clearly not possible to estimate \( p \) and \( p' \) (or \( P \)) if each individual is examined only once. The only function of the parameters which can be estimated from such data is essentially \( Pp + (1-P)p' \) - the probability that an individual chosen at random is declared NC - because the distribution depends only on this quantity.

2. ESTIMATION

If individuals are examined more than once, however, it is possible to estimate each of the three parameters. We will suppose that each of the \( n \) individuals in the random sample is examined on \( m \) independent occasions. If \( D_i \) denote the number of times the \( i \)-th individual is declared to be NC, it has the distribution

\[
\Pr[D_i = d_i] = \binom{m}{d_i} p^{d_i} (1-p)^{m-d_i} + (1-P) \binom{m}{d_i} p'_{i} (1-p')^{m-d_i}
\]

\((d_i = 0,1,...,m)\).

This is a mixture of two binomial distributions, with parameters \((m,p)\) and \((m,p')\) in proportions \( P \), \((1-P)\) respectively. The \( r \)-th factorial moment of each \( D_i \) is

\[
\mu_{(r)} = E[D_i^{(r)}] \\
= E[D_i (D_i-1) \ldots (D_i-r+1)] = m^{(r)} \{ Pp^r + (1-P)p'^r \}
\]

Estimating the parameters by making sample and population values of the first three factorial moments agree, we have
\[- 3 -
\]

\[ p^r + (1 - p)p'^r = F_r \quad (r=1,2,3) \]  

(3)

where \( F_r = \left(m(r)\right)^{-1} \sum_{i=1}^{n} D_i(r) \)

Solutions \( \tilde{p}, \tilde{p}' \) and \( \tilde{p} \) of (3) are given by Jones (1933) as follows

(a) \( \tilde{p}, \tilde{p}' \) are roots (in \( \theta \)) of the equation

\[ \theta^2 - A\theta + AF_1 - F_2 = 0 \]  

(4)

where \( A = \left(F_3 - F_1F_2\right)/\left(F_2 - F_1^2\right) \) (note that \( AF_1 - F_2 = \left(F_1F_3 - F_2^2\right)/\left(F_2 - F_1^2\right) \))

(b) \( \tilde{p} = \left(F_1\tilde{p}'\right)/\left(\tilde{p} - \tilde{p}'\right) \).

(5)

There is indeterminacy in the solution, since if \((\tilde{p}, \tilde{p}', \tilde{p})\) is a solution, so is \((\tilde{p}', \tilde{p}, 1-\tilde{p})\). We will adopt the convention of regarding the greater root of (4) as the estimator \( \tilde{p} \) of \( p \). It is reasonable to suppose that \( p \) is greater than \( p' \) - that is, the probability of declaring an individual to be NC if it is, indeed, NC is greater than if it is not. However, it must be remembered that even if this is so (i.e. \( p > p' \)), this does not ensure that \( \tilde{p} \) must exceed \( \tilde{p}' \).

3. Illustrative Example

For purposes of calculation, note that

\[ \sum_{i=1}^{n} D_i(r) = \sum_{j=1}^{m} N_j 3(r) \]  

(6)

where \( N_j \) = number of individuals declared NC on just \( j \) occasions among the \( m \) times examined.

Suppose we have \( n=50, m=3; N_0=43, N_1=1, N_2=1, N_3=5 \). Then
\[ F_r = (3^r \cdot 50)^{-1}(43.0^{r} + 1.1^r + 1.2^r + 5.3^r), \]

is

\[ F_1 = \frac{1}{150} (1+2+15) = \frac{3}{25}; \]

\[ F_2 = \frac{1}{300} (2+30) = \frac{8}{75}; \]

\[ F_3 = \frac{1}{300} \cdot 30 = \frac{1}{10}, \]

where \( A = \frac{327}{346} = 0.94508671 \) and \( \bar{p}, \bar{p}' \) are roots of \( \theta^2 - 0.945199 + 0.00674 = 0. \)

We find \( \bar{p} = 0.9379; \bar{p}' = 0.0072; \bar{P} = 0.1212. \)

In this case \( \bar{p}, \bar{p}' \) and \( \bar{P} \) are all between 0 and 1.

If they are not the method fails, though Blischke (1962, 1964) has suggested rules for this case.

4. Variances

Blischke (1962) has obtained the following asymptotic formulae for the variances of \( \bar{p}, \bar{p}' \) and \( \bar{P} : \)

\[
\text{var}(\bar{p}) = \frac{p(1-p)}{P} + 2\frac{(B_2 + B_2')}{nm} + \frac{6(B_3 + B_3')}{nm(m-1)(m-2)C_4} \quad (7.1)
\]

\[
\text{var}(\bar{p}') = \frac{p'(1-p')}{(1-P)nm} + 2\frac{(B_2 + 4B_2')}{nm(m-1)C_4^2} + \frac{6(B_3 + B_3')}{nm(m-1)(m-2)C_4^4} \quad (7.2)
\]

\[
\text{var}(\bar{P}) = \frac{P(1-P)}{n} + 2\frac{24(B_3 + B_3')}{nm(m-1)(p-p')^4} + \frac{24(B_3 + B_3')}{nm(m-1)(m-2)(p-p')^6} \quad (7.3)
\]
where \( B_h = P^h (1-p)^h \), \( B'_h = (1-P)p^h (1-p')^h \)

\[ C_h = P^2 (p-p')^h; \quad C'_h = (1-P)^2 (p-p')^h \]

When \( m \) is large, the first term of each expression usually gives quite good approximation, so that we may take

\[
\begin{align*}
\text{var}(\bar{p}) & \doteq \frac{p(p-1)}{P \cdot mn} \quad ; \quad \text{var}(\bar{p}') \doteq \frac{p'(1-p')}{(1-P)mn} \quad ; \quad \text{var}(\bar{P}) \doteq \frac{P(1-P)}{n} \, .
\end{align*}
\]

(8)

Note that the denominators are, for \( \text{var}(\bar{p}) \), the expected number of examinations of \( NC \) individuals; for \( \text{var}(\bar{p}') \), the expected numbers of examinations of conforming individuals; and for \( \text{var}(\bar{P}) \) the number of individuals in the sample.

Using the numerical values in the example of Section 3, and inserting the values \( \bar{p}, \bar{p}', \bar{P} \) for \( p, p', P \) respectively, we find (using (7.1)-(7.3))

\[
\begin{align*}
\text{var}(\bar{p}) & \doteq 0.004133 \\
\text{var}(\bar{p}') & \doteq 0.0000591 \\
\text{var}(\bar{P}) & \doteq 0.002170
\end{align*}
\]

The last two terms in the expressions on the right-hand sides of (7.1)-(7.3) are

\[
\begin{align*}
0.000885 \text{ and } 0.000044 & \text{ for } \text{var}(\bar{p}) \, ; \\
0.000004 \text{ and } 0.000001 & \text{ for } \text{var}(\bar{p}') \, ; \\
0.000036 \text{ and } 0.000003 & \text{ for } \text{var}(\bar{P}) .
\end{align*}
\]

So the use of (8) in this case, at least, would give quite good results,
even though \( m \) is only 3. (To the same order of approximation the three estimators are uncorrelated.)

5. Confidence Intervals

Blischke (1962) also showed that the asymptotic distributions of the estimators are normal. For \( P \), approximate 100(1-\( \alpha \))% confidence regions can be obtained from the inequality

\[
\frac{n(P-P')^2}{P(1-P)} < \chi^2_{\frac{1}{2} \alpha}
\]

where \( \Phi(-\chi_{\frac{1}{2} \alpha}) = \frac{1}{2} \alpha \) and \( \Phi(y) = (\sqrt{2\pi})^{-1} \int_{-\infty}^{y} e^{-\frac{u^2}{2}} du \). Taking \( \alpha = 0.05 \), so that \( \chi^2_{\frac{1}{2} 0.05} = 3.8416 \) the approximate 95% region for \( P \) is

\[
50(0.1212 - P)^2 < 3.8416 P(1-P)
\]

or equivalently

\[
53.8416 P^2 - 15.9616P + 0.73447 < 0
\]

that is \( 0.057 < P < 0.240 \).

Unfortunately we cannot use this method to obtain confidence regions for \( p \) and \( p' \). The corresponding region for \( p \) (using (8)) would be

\[
\frac{p \, mn(P-p)^2}{p(1-p)} < \chi^2_{\frac{1}{2} \alpha}
\]

which cannot be used because \( P \) is not known. We might replace \( P \) by \( \tilde{P} \). This would give an asymptotically correct region.

Using the value \( \tilde{P} = 0.1212 \), and taking \( \alpha = 0.05 \), as before, we get the (approximate) 95% confidence regions:


for $p$: $0.1212 \cdot 150(0.9379-p)^2 < 3.8416 \ p(1-p)$,  

whence $22.0216 \ p^2 - 37.9436p + 15.9922 < 0$  

leading to the interval $0.7350 < p < 0.9880$ ;

for $p'$: $18.18 \ (0.0072-p')^2 < 3.8416 \ p'(1-p')$  

whence $22.0216 \ p'^2 - 4.1034 \ p' + 0.00094 < 0$  

leading to the interval $0.00023 < p' < 0.0650$.

References

Blischke, W.R. (1962) Moment estimators for the parameters of a mixture of  


Generalizations". In Reliability in the Acquisitions Process (D.J.  

inspection, Statist. Hefte, 26, 19-29.


Jones, H.G. (1933) A note on the n-ages method. J. Inst. Actu., 64, 318- 
324.

Kotz, S. and Johnson, N.L. (1982) Errors in inspection and grading:  
Distributional aspects of screening and hierarchical screening. Commun.  

ACKNOWLEDGMENTS

Dr. Samuel Kotz's work was supported by the U.S. Office of Naval Research  
under Contract N 00014-84-K-301.

Professor Mao Shisong's work was carried out during a visit to the  
University of Maryland, College Park in 1985.
## Title

Estimation of Moments in a Model of Faulty Inspection

### Authors

Mao ShiSong, Samuel Kotz and Norman L. Johnson

### Type of Report

Technical

### Time Covered

FROM 7/85 TO 6/85

### Date of Report

June 1985

### Abstract

Methodology developed by Blischke (Ann. Math. Statist. 33(1962), 444-54) is applied to estimate the parameters in a model of faulty inspection, and to obtain approximate formula for the variances of these estimators.