THE INSTITUTE OF STATISTICS

THE CONSOLIDATED UNIVERSITY
OF NORTH CAROLINA

DETECTION OF FAULTY INSPECTION

by

Norman L. Johnson

Department of Statistics
University of North Carolina at Chapel Hill

and

Samuel Kotz

Department of Management Science & Statistics
University of Maryland

Institute of Statistics Mimeo Series #1506

August 1982
DEPARTMENT OF STATISTICS
Chapel Hill, North Carolina

DETECTION OF FAULTY INSPECTION

by

Norman L. Johnson

University of North Carolina
at Chapel Hill

Samuel Kotz

University of Maryland

ABSTRACT

Some heuristic tests for detection of existence of errors in inspection are proposed. Some of these tests are only effective if the sampling function ((sample size)/(lot size)) is rather large, and in all cases their application predicates special experiments to provide the requisite data. Feasibility of these experiments will vary according to specific circumstances.

Estimation of the probability of detection is discussed.

August 1982

1. INTRODUCTION

Recent papers (Johnson et al.)(1980), Johnson & Kotz (1981), 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, while possibly some non-defectives are described as 'defective' ("false positives"). Although these results are of interest, some more practical problems arise when it is desired to test whether inspection is faulty or to estimate the degree of imperfection. It is the object of this paper to discuss some aspects of these problems, keeping in mind possible practical constraints on availability of data.

2. SCOPE OF THE PROBLEM

It is clear that detection of faulty inspection will usually call for special investigation; the possible forms of such investigation can be limited by practical constraints. We will give a few possible modes of attack, but will deal here with the very simple case in which random samples of size n are taken from a lot of size N, containing X (unknown) defective items, with constant probability, p for each defective item, that it will indeed be classified as 'defective' on inspection, and with zero probability of false positives. Even with these simplifying assumptions, detection of faulty inspection will often be difficult; and there are clearly many possibilities of complication. For example it may well be that the class 'defectives' is not homogeneous; some may be more, and some less easily detected. (Sampling from such stratified populations is discussed in Johnson & Kotz (1981).)

If sampling is with replacement, it will not be possible to distinguish faulty from perfect inspection merely on the basis of a succession of values $Z_1, Z_2, \ldots$ of the total number (Z) of apparently defective items in routine samples of size n. This is because in this case each Z will have a binomial distribution with parameters n, pX/N and it will not be possible to separate the unknowns p and X, and so not possible to test the hypothesis p = 1. Some discrimination is possible if sampling is without replacement, but this will naturally be weak especially if the sampling fraction (f=n/N) or the proportion (θ=X/N) of defective items is small. This problem is discussed in Section 3. Further possibilities of specially designed experiments are described and discussed in Section 4. Estimation of p, and yet more possible special designs or discussed in Section 5.

3. AN EXPERIMENT TO TEST FOR FAULTY INSPECTION

We will suppose $Z_1, Z_2, ,, Z_m$ (the number of items declared defectives in successive samples) mutually independent. This would require the return to the population of all n items in each sample before the next sample is chosen (even though for each sample, selection is <u>without</u> replacement) but does not call for identification of previously chosen items. We will suppose this done, but note that it is evident that it will not always be possible - for example, if testing is destructive. In such cases we would have dependence among the Z's, arising from dependence among the corresponding numbers (Y) of actual defective items in the samples - in fact

$$\Pr[\bigcap_{j=1}^{m}(Y_j=y_j)\,|\,X] = \binom{X}{X-\sum_{j=1}^{m} y_j, y_1, \ldots, y_m}\binom{N-X}{N-X-mn+\sum_{j=1}^{m} y_j, n-y_1, \ldots, n-y_m}\Bigg/\binom{N}{N-mn, n, \ldots, n}$$

where $\binom{a}{b_0, b_1, \ldots, b_m} = a!/(\prod_{j=0}^{m} b_j!)$. We will not discuss this case further here.

The distribution of each Z is the hypergeometric-binomial

$$\Pr[Z=z|X,p] = \binom{N}{n}^{-1} \sum_{y>z} \binom{X}{y}\binom{N-X}{n-y}\binom{y}{z} p^z(1-p)^{y-z} \qquad (1)$$

$$(\max(0,n-N+X) \le z \le \min(n,X))$$

(Johnson et al. (1980)). The mean (μ) and variance (σ^2) of Z are

$$\mu = np\, X/N (= np\theta = Npf\theta) \qquad (2.1)$$

$$\sigma^2 = n\{\frac{pX}{N}(1-\frac{pX}{N}) - \frac{n-1}{N-1}\frac{X}{N}(1-\frac{X}{N})p^2\}$$

$$= n\{p\theta(1-p\theta) - \frac{n-1}{N-1}p^2\theta(1-\theta)\}$$

$$= \frac{N-n}{N-1}\mu(1-n^{-1}\mu) + \frac{n-1}{N-1}(1-p)\mu \quad \left[\doteq (1-f)\mu(1-n^{-1}\mu) + f(1-p)\mu\right] \qquad (2.2)$$

More generally the r-th factorial moment of Z is

$$\mu_{(r)}(Z) = E[Z^{(r)}] = E[Z(Z-1)..(Z-r+1)] = n^{(r)}X^{(r)}p^r/N^{(r)} \qquad (2.3)$$

Formally our problem is that of testing the hypothesis H_o:$p = 1$, with X as nuisance parameter. It is clear that H_o cannot be tested on the basis of a single observed Z value. Even with a set of independent values $Z_1, Z_2, \ldots Z_m$, (obtainable if each sample of m items is returned to the lot after inspection) construction of a test, which is optimal in some reasonable way, presents technical difficulties.

Since μ is estimated unbiasedly by $\bar{Z} = m^{-1}\sum_{i=1}^{m} Z_i$, and σ^2 by $S^2 = (m-1)^{-1}\sum_{i=1}^{m}(Z_i-\bar{Z})^2$, it is natural to seek for some function of μ and σ^2 which depends only on p (and not on X) and then replace μ by $\bar{Z}$ and σ^2 by S^2. From (2.2)

$$\Omega = \sigma^2\mu^{-1} + n^{-1}(N-1)^{-1}(N-n)\mu = 1-(N-1)^{-1}(n-1)p \qquad (2.4)$$

is such a function and so we consider using

$$W = S^2\bar{Z}^{-1} + n^{-1}(N-1)^{-1}(N-n)\bar{Z} \qquad (3)$$

as a test criterion. The hypothesis $H_o(p=1)$ corresponds to $\Omega = (N-1)^{-1}(N-n)$; for alternative hypotheses ($p < 1$), $\Omega > (N-1)^{-1}(N-n)$.

Using standard approximate ("statistical differentials") formulae, we have

$$E[W|X,p] = E[S^2\bar{Z}^{-1}|X,p] + N^{-1}(N-1)^{-1}(N-n)p\,X$$

$$\doteq \frac{\sigma^2}{\mu}\{1 - \frac{\text{cov}(S^2,\bar{Z}|X,p)}{\sigma^2\mu} + \frac{\text{var}(\bar{Z}|X,p)}{\mu^2}\} + n^{-1}(N-1)^{-1}(N-n)\mu$$

$$\doteq \frac{\sigma^2}{\mu}[1 - m^{-1}\{\sqrt{\beta_1}\frac{\sigma}{\mu} - (\frac{\sigma}{\mu})^2\}] + n^{-1}(N-1)^{-1}(N-n)\mu \qquad \text{(see Neyman (1926))}$$

$$\doteq \Omega - m^{-1}\sigma(\frac{\sigma}{\mu})^2(\sqrt{\beta_1}-\frac{\sigma}{\mu}) \tag{4.1}$$

Also

$$\text{var}(W|X,p) = \text{var}(S^2\bar{Z}^{-1}|X,p)+2n^{-1}(N-1)^{-1}(N-n)\text{cov}(S^2\bar{Z}^{-1},\bar{Z}|X,p)$$
$$+ n^{-2}(N-1)^{-2}(N-n)^2\text{var}(\bar{Z}|X,p)$$

$$\doteq \frac{\sigma^4}{\mu^2}[\frac{\text{var}(S^2|X,p)}{\sigma^4} - \frac{2\text{cov}(S^2,\bar{Z}|X,p)}{\sigma^2\mu} + \frac{\text{var}(\bar{Z}|X,p)}{\mu^2}]$$
$$+ 2n^{-1}(N-1)^{-1}(N-n)[\sigma^2 - \frac{\sigma^2}{\mu}\cdot\mu\{1-m^{-1}\mu(\sqrt{\beta_1} - \frac{\sigma}{\mu})\}]$$
$$+ m^{-1}n^{-2}(N-1)^{-2}(N-n)^2\sigma^2$$

$$\doteq \frac{\sigma^4}{m\mu^2}[\beta_2 - \frac{m-3}{m-1} - 2\frac{\sigma}{\mu}\sqrt{\beta_1} + (\frac{\sigma}{\mu})^2 + 2n^{-1}(N-1)^{-1}(N-n)(\frac{\mu}{\sigma}\sqrt{\beta_1}-1)$$
$$+ n^{-2}(N-1)^{-2}(N-n)^2(\frac{\mu}{\sigma})^2] \tag{4.2}$$

where $\sqrt{\beta_1}$, β_2 are the moment ratios

$$\sqrt{\beta_1} = \sigma^{-3}E[(Z-\mu)^3|X,p];\ \beta_2 = \sigma^{-4}E[(Z-\mu)^4|X,p] \tag{5}$$

of the distribution (1).

When the null hypothesis ($p = 1$) is valid, this distribution is hypergeometric with parameters (n,X,N); the appropriate formulae for $\sqrt{\beta_1}$ and β_2 are given in Johnson & Kotz (1969, p. 144). If N is large, so that $(N-an)/(N+b) \doteq 1 - af$, then

$$\sqrt{\beta_1} \doteq \frac{(1-2f)(1-2\theta)}{\sqrt{\{n'\theta(1-\theta)\}}}\ ;\ \beta_2 \doteq 3 + \frac{1}{n'\theta(1-\theta)}[1-6\{f(1-f)+\theta(1-\theta)\}] \tag{6}$$

where $n' = n(1-f)$. We also have

$$\mu = n'\theta \tag{7.1}$$
$$\sigma^2 = n'\theta(1-\theta) \tag{7.2}$$

Inserting these values in (4.1) and (4.2) with $p = 1$, we obtain

$$E[W|X,1] \doteq \frac{N-n}{N-1} + \frac{(1-\theta)(2f+\theta-4f\theta)}{mn'\theta} \tag{8.1}$$

$$\text{var}(W|X,1) \doteq \frac{(1-\theta)^2}{m}\left[\frac{2m}{m-1} + \frac{1}{n'\theta(1-\theta)}\{f^2(f\theta+2\theta-2)^2 - 2f(1-f) - 2\theta(1-\theta)\right.$$

$$\left. - 16f(1-f)\theta(1-\theta)\}\right]$$

$$= \frac{2(1-\theta)^2}{m-1} + \frac{1-\theta}{n'm\theta}\, g(f,\theta) \tag{8.2}$$

where $g(f,\theta)$ is a number of magnitude about 1 (for example if $f = \theta = \frac{1}{2}$, $g(f,\theta) = \frac{1}{4}(\frac{1}{4} + \frac{2}{2} - 2)^2 - \frac{1}{2} - \frac{1}{2} - (16 \times \frac{1}{16}) = \frac{9}{64} - 2 = -1\frac{55}{64}$).

4. OTHER APPROACHES

As we have already noted the most straightforward way of testing the hypothesis $p = 1$ would be by inspection of items known to be defective. Standard methods could then be used - in this case the hypothesis would be rejected as soon as any item was 'found' to be nondefective. The level of significance of the test would be zero, and its power function $(1-p^m)$ where m is the number of items to be inspected.

Such an inquiry may not be possible (for example when determination of defectiveness entails destruction) but we may approach it by repeated inspection of the same item (not knowing whether it is defective or not). If any two inspections give different results, then we know that $p \neq 1$ (since we are here assuming that a nondefective item will always be classified correctly). Of course, if the item is really nondefective this test will have no discriminating power, since every inspection will result in a correct decision.

Sometimes it may be possible to repeat inspection of the same sample (of size n) m times, but to be able to record only the total numbers $(T_1, T_2, \ldots, T_m)$ of items deemed 'defective' at each inspection. This increases the chance of including at least one truly defective item among those inspected.

Here, again, if any two T's are unequal, then we know that H_0 is not true (i.e. $p < 1$) since decision in regard to at least one item must differ in the two corresponding inspections. If there are really y defective items among the n subjected to inspection, the probability that departure of p from 1 will be detected, as a consequence of at least two of the T's differing is

$$1 - \sum_{t=0}^{y} \{\binom{y}{t} p^t (1-p)^{y-t}\}^m .$$

The overall probability, if the sample is chosen without replacement from a population of size N containing just X defectives is

$$1 - E[\sum_{t=0}^{Y} \{\binom{Y}{t} p^t (1-p)^{Y-t}\}^m]$$

where Y has a hypergeometric distribution with parameters (n,X,N).

5. ESTIMATION OF THE PROBABILITY OF DETECTION (p)

So far we have considered only testing for the existence of errors in inspection which lead to nondetection of defective items. If the existence of such errors is established, it is natural to attempt to estimate p. With data of the kind used in the test statistic W (Section 3) a natural estimator would be (in view of (4.1)) $(n-1)(N-1)^{-1}(1-W)$, possibly with a bias correction. For reasonable accuracy however, we need data of the kind described towards the end of Section 4.

If we were in the fortunate position of having a number of items known to be defective we could obtain a simple estimator (of p) by testing them repeatedly and estimating p by the proportion of times a 'defective' decision is obtained.

If this is not the case at the beginning of the investigation we might, however, be in a position to exploit the fact that an item declared 'defective' at any time must (according to our assumptions) be defective. Denoting by N_j the number of items declared defective just j times in m trials (so that $N_0 + N_1 + \ldots + N_m = n$, and if $p = 1$, $N_0 = n-y$, $N_m = y$) a plausible but specious argument might run as follows:

> "For each item, we discard the first 'defective' decision and observe the proportion of defective decision in the remaining $\sum_{j=1}^{m} N_j$ sets of (m-1) trials. Since these are independent, the total number of defectives in the trials has a binomial $((m-1)\sum_{j=1}^{m} N_j, p)$ distribution, and our estimate of p is unbiased, with an easily computed standard deviation."

(It is not difficult to see that this will produce a negatively biased estimator of p, because in all the trials which are thrown away a decision of 'defective' is reached. In fact the estimation is

$$\frac{1}{(m-1)\sum_{j=1}^{m} N_j}\left[\sum_{j=1}^{m} (j-1)N_j\right] = \frac{1}{m-1}\left[\frac{\sum_{j=1}^{m} jN_j}{\sum_{j=1}^{m} N_j} - 1\right] \tag{9}$$

and its expected value is

$$\left\{\frac{1}{m-1}\;\frac{mp}{1-(1-p)^m} - 1\right\} \cdot\;) \tag{10}$$

We could take notice of only those trials following the first 'defective' decision; although this will not use all the information available, it does lead to simple formulae. The observed proportion of 'defective' decisions is no an unbiased estimator of p; the (conditional) distribution of the number of 'defective decision counted is binomial with parameters (total number of inspections included in count, p).

Suppose, now, that we have m repeated inspections of a set of n items (among which an unknown number y are defective) and have been able to record the results of individual inspections (and not just total number of decisions of 'defective' for each inspection of the set of n items). The likelihood function is

$$\binom{y}{y-n+N_0, N_1, \ldots, N_m}(1-p)^{m(y-n+N_0)} \prod_{j=1}^{m} \left\{\binom{m}{j}p^j(1-p)^{m-j}\right\}^{N_j}$$

$$= \binom{y}{y-n+N_0, N_1, \ldots, N_m}\left\{\prod_{j=1}^{m} \binom{m}{j}\right\} p^{\sum_1^m jN_j}(1-p)^{my-\sum_1^m jN_j} \qquad (11)$$

$$(n-N_0 \leq y \leq n)$$

where $N_0 = n - \sum_{j=1}^{m} N_j$ is the number of items which are not declared defective is any of the m inspections. Note that $(N_0, \sum_{j=1}^{m} jN_j)$ is a sufficient statistic for (y,p); $\sum jN_n$ is the total number of 'defective' decisions. If y were known, the maximum likelihood estimator of p would be

$$\hat{p}(y) = (my)^{-1}\sum_{j=1}^{m} jN_j \qquad (12)$$

The corresponding maximized log likelihood would be

$$\log \hat{L}(y) = K + \sum_{i=0}^{n-N_0+1} \log(y-i) + (my-\sum_{j=1}^{m} jN_j)\log(my-\sum_{j=1}^{m} jN_j) - my \log my \qquad (13)$$

where K does not depend on y. We then seek to minimize (13) with respect to y, subject to $n \geq y \geq n - N_0$. Note that we are not primarily interested in the value of y itself, but we need the value, $\hat{y}$, maximizing (13) to calculate the maximum likelihood estimator, $\hat{p}(\hat{y})$, of p.

A useful practical method is obtained by noting that for the $\sum_{j=1}^{m} N_j$ items which we know to be defective, the numbers of times each item is declared 'defective' can be regarded as observed values of independent random variables each having a binomial (m,p) distribution, truncated by omission of zero values. Equating sample means and expected values gives the equation

$$\tilde{p}\{1-(1-\tilde{p})^m\}^{-1} = m^{-1}(n-N_0)^{-1}\sum_{j=1}^{m} jN_j \tag{14}$$

for an estimator of p. This estimator is, in fact, the conditional maximum likelihood estimator of p, given N_0.

If we do require to get a estimator of y, we note that, with $p = \tilde{p}$, $E[\sum_{j=1}^{m} N_j = n - N_0 | y,\tilde{p}] = y\{1-(1-\tilde{p})^m\}$.

Replacing expected by actual values, we get the the estimator

$$\tilde{y} = \min(n, \left[(n-N_0)\{1-(1-\tilde{p})^m\}^{-1}\right]) = \min(n, \left[(m\tilde{p})^{-1}\sum_{j=1}^{m} jN_j\right]) \tag{15}$$

where [ ] denotes 'nearest integer to'.

As a numerical example suppose we test each of 50 (=n) items three (= m) times and obtain

$$N_0 = 43;\ N_1 = 1;\ N_2 = 1;\ N_3 = 5$$

so that

$$\sum_{j=1}^{3} jN_j = 1+2+15 = 18.$$ Equation (14) gives

$$\tilde{p}\{3\tilde{p} - 3\tilde{p}^2 + \tilde{p}^3\}^{-1} = (3 \times 7)^{-1}18 = 6/7$$

where $\tilde{p}^2 - 3\tilde{p} + \frac{11}{6} = 0$

leading to $\tilde{p} = \underline{0.8545}$

From (15) , $\tilde{y} = \min\left(50, \frac{18}{3 \times 0.8545}\right) = 7$.

Note that the same values of $\tilde{p}$ and $\tilde{y}$ would be obtained, whatever the value of n ($\geq$ 7).

ACKNOWLEDGEMENT

Samuel Kotz's work was supported by the U.S. Office of Naval Research under Contract N00014-81-K-0301. Norman L. Johnson's work was supported by the National Science Foundation under Grant MCS-8021704.

REFERENCES

Johnson, N.L. and Kotz, S. (1969) *Distributions in Statistics - Discrete Distributions*, Wiley, New York.

Johnson, N.L. and Kotz, S. (1981) "Faulty inspection distributions - some generalizations". To appear in *Proceedings of the ONR/ARO Reliability Workshop*, April 1981 (D.J. de Prieto and R.L. Launer, Eds.) Marcel Dekker, New York.

Johnson, N.L., Kotz, S. and Sorkin, H.L. (1980) "Faulty inspection distributions", *Commun. Statist.* **A9**, 917-922.

Kotz, S. and Johnson, N.L. (1982) "Errors in inspection and grading: Distributional aspects of screening and hierarchal screening", *Commun. Statist.*, **A11**,

Neyman, J. (1926) "On the correlation of the mean and variance in samples drawn from an 'infinite' population," *Biometrika*, **18**, 401-413.