

SOME MULTIVARIATE DISTRIBUTIONS ARISING
IN FAULTY SAMPLING INSPECTION

Norman L. Johnson

Samuel Kotz

University of North Carolina
at Chapel Hill

and

University of Maryland
at College Park

An Invited Paper for the Sixth International
Symposium on Multivariate Analysis

University of Pittsburgh, PA
July 1983

January 1983

ABSTRACT

In the present paper we extend in two ways some results presented in Kotz and Johnson (Comm. in Statistics (1982), A11) relating to the study of distributional aspects of effects of errors in inspection sampling:

(1) Multistage sampling with k successive samples involving the possibility of two types of errors in inspection (classifying a defective individual as non-defective, or a non-defective as defective); (2) Single-stage sampling considering several types of defects of which only one is tested on inspection. Both (1) and (2) lead to novel multivariate distributions. Their structural properties are analysed in some detail and some applications, in particular those in quality control are discussed.

1. INTRODUCTION

We have recently studied effects of inaccuracies in inspection on the properties of acceptance sampling procedures (Johnson & Kotz (1981), Kotz & Johnson (1982a,b)). In particular we have considered two-stage sampling, wherein up to two successive samples of sizes n_1, n_2 may be taken from a lot of size N , containing D defective items. (The second sample is taken only if the number (Z_1) of items found to be defective is within certain limits defined by the sampling schemes.) It was supposed that inspection is not perfect, so that some defective items may not be noticed as such, while some nondefective items may be classified as 'defective'.

Similar investigations have been described in quality control literature by Hoag et al. (1975) Dorris and Foote (1978) and Rahali and Foote (1982) (see also Armstrong (1982)). In the present paper we shall extend this work along two directions, each of which introduces apparently novel multivariate distributions.

First we shall generalize our result to multistage sampling with k successive samples of sizes n_1, n_2, \dots, n_k and suppose that for the i -th sample

p_i = probability that a defective item is correctly classified

p_i' = probability that a nondefective item is classified as 'defective'.

Secondly, returning to single-stage sampling we shall consider the case of two types of defects one of which is relatively easy to detect, and investigate the distribution of the second kind of defect when selection is on

the basis of the first kind. As before we will suppose that p_1 denotes the probability of correct classification of a defective item and p_1' the probability of classification of a non-defective item as 'defective'.

2. MULTISTAGE SAMPLING: ANALYSIS

We will be interested in the joint distribution of Z_1, Z_2, \dots, Z_k , the numbers of items classified as defective as a result of inspection of the sequence of k samples. This distribution seems to be new, and presents some features of theoretical interest, as well as having the possible practical applications we have indicated. Chief among the latter is the possible calculation of acceptance probabilities for multi-stage (or even sequential) sampling schemes under imperfect inspection. Some practical comments, in the two-stage case are given in Kotz & Johnson (1982b).

Let Y_1, Y_2, \dots, Y_k denote the actual numbers of defectives in the 1st, 2nd, ..., k -th samples respectively. The joint distribution of $\underline{Y} = (Y_1, \dots, Y_k)$ is a multivariate hypergeometric with

$$\Pr\left[\prod_{i=1}^k (Y_i = y_i)\right] = \binom{N}{D}^{-1} \left\{ \prod_{i=1}^k \binom{n_i}{y_i} \right\} \binom{N - \sum y_i}{D - \sum y_i} \quad (1)$$

$$(0 \leq y_i < n_i (i=1, \dots, k); D - N + \sum_{i=1}^k n_i < \sum_{i=1}^k y_i).$$

Symbolically, we write

$$\underline{Y} \sim \text{Mult.Hypg}_k(\underline{n}; D, N).$$

We note that

$$E\left[\prod_{i=1}^k \binom{r_i}{Y_i} \binom{s_i}{n_i - Y_i}\right] = \frac{\binom{\sum r_i}{D} \binom{\sum s_i}{N-D}}{\prod_{i=1}^k \binom{r_i + s_i}{n_i}} \quad (2)$$

where $a^{(b)} = a(a-1)\dots(a-b+1)$ is the b -th descending factorial of a .

Conditionally on \underline{Y} , the Z 's are mutually independent, with Z_i distributed as

the sum of two independent binomial variables with parameters (Y_i, p_i) and $(n_i - Y_i, p_i')$ respectively.

$$\begin{aligned} \Pr\left[\prod_{i=1}^k (Z_i = z_i) \mid \underline{Y} = \underline{y}\right] &= \prod_{i=1}^k \left[\sum_{h=0}^{z_i} \binom{y_i}{h} \binom{n_i - y_i}{z_i - h} p_i^h p_i'^{z_i - h} (1 - p_i)^{y_i - h} (1 - p_i')^{n_i - y_i - z_i + h} \right] \\ &= \prod_{i=1}^k b(z_i; y_i, p_i; n_i - y_i, p_i') \end{aligned} \quad (3)$$

where $b(z_i; y_i, p_i; n_i - y_i, p_i')$ is the probability function for the convolution of the two binomial distributions with parameters (y_i, p_i) and $n_i - y_i, p_i'$ respectively.

The unconditional distribution of \underline{Z} is

$$\Pr[\underline{Z} = \underline{z}] = \binom{N}{D}^{-1} \sum_{y_1} \dots \sum_{y_k} \left\{ \binom{N - \sum n_i}{D - \sum y_i} \prod_{i=1}^k \binom{n_i}{y_i} \right\} \prod_{i=1}^k b(z_i; y_i, p_i; n_i - y_i, p_i') \quad (4)$$

The limits for \underline{y} are as in (1).

Symbolically

$$\underline{Z} \sim \left(\begin{array}{cc} \text{Bin}(Y_1, p_1) * \text{Bin}(n_1 - Y_1, p_1') \\ \vdots \quad \quad \quad \vdots \\ \text{Bin}(Y_k, p_k) * \text{Bin}(n_k - Y_k, p_k') \end{array} \right) \underset{\underline{Y}}{\Lambda} \text{Mult.Hypg}_k (n: D, N) \quad (4a)$$

where * stands for convolution and Λ is the compounding operator (as defined, for example, in Johnson & Kotz (1969, p. 184)).

This distribution might be called "Multivariate Hypergeometric-Convolutated Binomial(s)".

The conditional (on \underline{y}) r_i -th factorial moment of Z_i is

$$E[Z_i^{(r_i)} \mid \underline{y}] = \sum_{h=0}^{r_i} \binom{r_i}{h} p_i^h p_i'^{r_i - h} \binom{y_i}{h} \binom{n_i - y_i}{r_i - h} (1 - p_i)^{y_i - h} (1 - p_i')^{n_i - y_i - r_i + h}$$

The unconditional $r = (r_1, \dots, r_k)$ -th factorial moment of Z is

$$\begin{aligned} \mu_r(Z) &= E\left[\prod_{i=1}^k Z_i^{(r_i)}\right] = E_Y E\left[\prod_{i=1}^k Z_i^{(r_i)} \mid Y\right] = E_Y \left[\prod_{i=1}^k E[Z_i^{(r_i)} \mid Y]\right] \\ &= E\left[\prod_{i=1}^k \left\{ \sum_{h=0}^{r_i} \binom{r_i}{h} p_i^h p_i'^{r_i-h} Y_i^{(h)} (n_i - Y_i)^{(r_i-h)} \right\}\right] \quad (6) \\ &= \sum_{\ell_1=0}^{r_1} \dots \sum_{\ell_k=0}^{r_k} \left\{ \prod_{i=1}^k \binom{r_i}{\ell_i} p_i^{\ell_i} p_i'^{r_i-\ell_i} E[Y_i^{(\ell_i)} (n_i - Y_i)^{(r_i-\ell_i)}] \right\} \\ &= \frac{\prod_{i=1}^k n_i^{(r_i)}}{N^{(\sum r_i)}} \sum_{\ell_1=0}^{r_1} \dots \sum_{\ell_k=0}^{r_k} \left\{ \prod_{i=1}^k \binom{r_i}{\ell_i} p_i^{\ell_i} p_i'^{r_i-\ell_i} \right\} D^{(\sum \ell_i)} (N-D)^{(\sum (r_i - \ell_i))} \\ &= \frac{\prod_{i=1}^k n_i^{(r_i)}}{N^{(\sum r_i)}} \sum_{\ell=0}^{\sum r_i} D^{(\ell)} (N-D)^{(\sum r_i - \ell)} g_\ell \quad (6') \end{aligned}$$

where g_ℓ is the coefficient of x^ℓ in $\prod_{i=1}^k (p_i' + p_i x)^{r_i}$. (Note that

(i) if sampling were with replacement, (6) would be

$$\mu_r(Z) = \frac{\prod_{i=1}^k n_i^{(r_i)}}{\prod_{i=1}^k \{D p_i + (N-D) p_i'\}^{r_i}} / N^{\sum r_i};$$

(ii) if $p_i = p$ and $p_i' = p'$ for all i -corresponding to constant quality of inspection throughout - then

$$g_\ell = \binom{\sum r_i}{\ell} p^\ell p'^{\sum r_i - \ell}$$

In particular

$$E[Z_i] = n_i [DN^{-1} p_i + (1-DN^{-1}) p_i'] = n_i \bar{p}_i \quad (7.1)$$

where $\bar{p}_i = DN^{-1} p_i + (1-DN^{-1}) p_i'$ is the probability that an individual chosen at random in the i -th sample will be classified as defective (whether it really is so, or not). Also

$$\text{var}(Z_i) = n_i \bar{p}_i (1 - \bar{p}_i) - n_i (n_i - 1) (N - 1)^{-1} DN^{-1} (1 - DN^{-1}) (p_i - p_i')^2 \quad (7.2)$$

and

$$\text{cov}(Z_i, Z_j) = -n_i n_j (N - 1)^{-1} DN^{-1} (1 - DN^{-1}) (p_i - p_i') (p_j - p_j') \quad (8)$$

If $p_i > p_i'$ and $p_j > p_j'$ as one would hope, if inspection is to be of any use at all, the covariance is negative, as might be expected, since the covariance of Y_i and Y_j is negative. If $p_i = p_i'$ or $p_j = p_j'$, so that it is irrelevant whether an item in the corresponding sample(s) is defective or not, then the covariance is zero - in fact the corresponding Z (or Z 's) is independent of all other Z 's.

It is easy to write down the conditional distribution of Z_i , given Z_j , but it is rather complicated in form. We can, however, derive the regression function in the following way.

We have
$$E[Z_i | Y_i] = Y_i p_i + (n_i - Y_i) p_i'$$

so
$$E[Z_i | Z_j] = p_i E[Y_i | Z_j] + p_i' (n_i - E[Y_i | Z_j]) \quad (9)$$

Now
$$E[Y_i | Y_j] = n_i (D - Y_j) / (N - n_j)$$

so
$$E[Y_i | Z_j] = n_i (N - n_j)^{-1} (D - E[Y_j | Z_j]) \quad (10)$$

In order to evaluate $E[Y_j | Z_j]$ note that, for the j -th sample,

$$\text{Pr}[\text{item is defective} | \text{item classified as defective}] = DN^{-1} p_j / \bar{p}_j$$

and

$$\text{Pr}[\text{item is defective} | \text{item classified as nondefective}] = DN^{-1} (1 - p_j) / (1 - \bar{p}_j)$$

so
$$E[Y_j | Z_j] = DN^{-1} [Z_j (p_j / \bar{p}_j) + (n_j - Z_j) \{(1 - p_j) / (1 - \bar{p}_j)\}] \quad (11)$$

From (9), (10) and (11) we obtain

$$\begin{aligned}
 E[Z_i | Z_j] &= n_i p_i' + (p_i - p_i') E[Y_i | Z_j] \\
 &= n_i p_i' + n_i (p_i - p_i') (N - n_j)^{-1} (D - E[Y_j | Z_j]) \\
 &= n_i p_i' + n_i (p_i - p_i') D N^{-1} \left\{ 1 - \frac{(1 - D N^{-1})(p_j - p_j')}{(N - n_j) \bar{p}_j (1 - \bar{p}_j)} \cdot (Z_j - n_j \bar{p}_j) \right\} \quad (12)
 \end{aligned}$$

The regression is linear, and the sign of the regression coefficient is opposite to that of $(p_i - p_i')(p_j - p_j')$ agreeing with the sign of $\text{cov}(Z_i, Z_j)$ in (8).

3. MULTISTAGE SAMPLING: A SPECIAL CASE

If we take $n_1 = n_2 = \dots = n_k = 1$ we have the first k stages of a fully sequential sampling procedure. In this situation the only possible values of each of the Y 's and Z 's are 0 and 1. Formula (3) becomes

$$\Pr[\underline{Z} = \underline{z}] = \binom{N}{D}^{-1} \sum_{y_1} \dots \sum_{y_k} \binom{N-k}{D - \sum y_i} \prod_{i=1}^k \{ y_i p_i^{z_i} (1-p_i)^{1-z_i} + (1-y_i) p_i' z_i (1-p_i')^{1-z_i} \} \quad (13)$$

Collecting together terms with the same value of $\sum y_i = y$ (corresponding to the total number of defective items selected) we get

$$\Pr[\underline{Z} = \underline{z}] = \binom{N}{D}^{-1} \sum_y \binom{N-k}{D-y} \sum_{\underline{\alpha}} \sum_{\underline{\beta}} \prod_{i=1}^k p_i^{\alpha_i} p_i' z_i^{\alpha_i'} (1-p_i)^{\beta_i} (1-p_i')^{\beta_i'} \quad (13)'$$

where summation with respect to $\underline{\alpha} = (\alpha_1, \dots, \alpha_k)$ and $\underline{\beta} = (\beta_1, \dots, \beta_k)$ is constrained by

$$\sum_{i=1}^k (\alpha_i + \beta_i) = y, \text{ and } \sum_{i=1}^k (\alpha_i + \alpha_i') = \sum_{i=1}^k z_i = z$$

(corresponding to the total number of items classified as defectives) and for each i , one of $(\alpha_i, \alpha_i', \beta_i, \beta_i')$ is 1, the other three are each zero.

Yet another way of writing (13) is

$$\Pr[Z=z] = \binom{N}{D}^{-1} \sum_y \binom{N-k}{D-y} \cdot [\text{coefficient of } x^z u^y \text{ in} \\ \prod_{j=1}^k \{p_j x u + p'_j x + (1-p_j) u + 1 - p'_j\}] \quad (13)''$$

It may be noted that this shows that $\sum_{i=1}^k Z_i$ is a sufficient statistic (for D, supposing N, p's and p's are known). If $p_i = p$ and $p'_i = p'$ for all i, we have the coefficient of $x^z u^y$ in $\{pxu + p'x + (1-p)u + (1-p')\}^k$ on the right hand side of (13)'' .

4. ASSOCIATED DEFECTS: ANALYSIS

We now suppose that there are two types of defects - (1) and (2) - and that in a population of size N there are D_{gh} individuals with g type (1), and h type (2) defects (g,h = 0,1). (Of course $D_{00} + D_{01} + D_{10} + D_{11} = N$.) For example (1) might represent surface irregularity, with (2) corresponding to internal flaws in material, such as metal bars or plates. In many situations the different types of defect correspond to different modes of failure.

A random sample of size n is taken (without replacement) from the population; and each of the chosen individuals is examined for presence of type (1) defect. We are interested in the number (Z_1) of individuals classified as possessing defect (1); and among these Z_1 individuals, the number (Z_1^*) actually possessing defect (1) and the number (Z_2^*) possessing defect (2).

The distribution of Z_1 is given in Section 2, and also in Kotz & Johnson (1982b). The distribution of Z_1^* is the hypergeometric-binomial

$$\text{Binomial}(Y, p_1) \underset{Y}{\wedge} \text{Hypg.}(n; D_{10} + D_{11}; N) \quad (14)$$

The probability that the sample will contain y_{gh} individuals with g type (1) defects and h type (2) defects is

$$\Pr \left[\prod_{g=0}^1 \prod_{h=0}^1 (Y_{gh} = y_{gh}) \right] = \left[\prod_{g=0}^1 \prod_{h=0}^1 \binom{D_{gh}}{y_{gh}} \right] / \binom{N}{n} \quad (y_{00} + y_{01} + y_{10} + y_{11} = n) \quad (15)$$

(A multivariate hypergeometric $(n; \underline{D}, N)$ distribution -cf (4a) where \underline{n} is a vector but D is not.) For the $\underline{\alpha} = (\alpha_{00}, \alpha_{01}, \alpha_{10}, \alpha_{11})$ -th joint factorial moment of \underline{Y} we have the expression

$$\mu_{(\underline{\alpha})}(\underline{y}) = n^{(\underline{\alpha})} \left[\prod_{g=0}^1 \prod_{h=0}^1 \binom{D_{gh}}{y_{gh}} \right] / N^{(\underline{\alpha})} \quad \text{where } \alpha = \alpha_{00} + \alpha_{01} + \alpha_{10} + \alpha_{11} \quad (16)$$

Given $y = (y_{00}, y_{01}, y_{10}, y_{11})$, Z_1 is distributed as the sum of two independent binomial variables, with parameters $(y_{10} + y_{11}, p_1)$ and $(y_{00} + y_{01}, p_1')$ respectively. The first of these variables is, in fact Z_1^* . Similarly, Z_2^* is distributed as the sum of two independent binomial variables with parameters (y_{11}, p_1) and (y_{01}, p_1') respectively. Introducing four independent binomial variables W_{gh} with parameters $(y_{gh}, gp_1 + (1-g)p_1')$ for $g, h = 0, 1$ then conditionally on $\underline{Y} = \underline{y}$

$$Z_1 = W_{00} + W_{01} + W_{10} + W_{11} \quad (17.1)$$

$$Z_1^* = W_{10} + W_{11} \quad (17.2)$$

$$Z_2^* = W_{01} + W_{11} \quad (17.3)$$

We have

$$E[Z_1 | \underline{y}] = (y_{10} + y_{11})p_1 + (y_{00} + y_{01})p_1'; \quad \text{var}(Z_1 | \underline{y}) = (y_{10} + y_{11})p_1(1-p_1) + (y_{01} + y_{00})p_1'(1-p_1') \quad (18.1)$$

$$E[Z_2^* | \underline{y}] = y_{11}p_1 + y_{01}p_1'; \quad \text{var}(Z_2^* | \underline{y}) = y_{11}p_1(1-p_1) + y_{01}p_1'(1-p_1') \quad (18.2)$$

$$\text{cov}(Z_1, Z_2^* | \underline{y}) = \text{var}(W_{10} + W_{11} | \underline{y}) = \text{var}(Z_2^* | \underline{y}) \quad (18.3)$$

$$\text{corr}(Z_1, Z_2^* | \underline{y}) = \{ \text{var}(Z_2^* | \underline{y}) / \text{var}(Z_1 | \underline{y}) \}^{1/2} \quad (18.4)$$

$$\text{cov}(Z_1^*, Z_2^* | \underline{y}) = \text{var}(W_{11} | \underline{y}) = y_{11}p_1(1-p_1) \quad (18.5)$$

$$\text{corr}(Z_1^*, Z_2^* | y) = \left[\frac{y_{11}^2 p_1 (1-p_1)}{(y_{10} + y_{11}) \{y_{11} p_1 (1-p_1) + y_{01} p_1' (1-p_1')\}} \right]^{\frac{1}{2}} \quad (18.6)$$

The distribution of Z_2^* is the multivariate hypergeometric-convoluted binomial

$$\{\text{Binomial}(Y_{11}, p_1) * \text{Binomial}(Y_{10}, p_1')\} \wedge_{Y_{11}, Y_{10}} \text{Mult. Hypg}(n; D_{11}, D_{10}; N) \quad (19)$$

Moments can be obtained similarly as in Section 2.

For an individual chosen at random from the population

$$\begin{aligned} \text{Pr}[\text{classified as having defect (1)}] &= N^{-1} \{(D_{01} + D_{00}) p_1' + (D_{10} + D_{11}) p_1\} \\ &= N^{-1} (D_0 p_1' + D_1 p_1) = \bar{p}_1, \text{ say} \end{aligned} \quad (19.1)$$

($D_1 = D_{10} + D_{11}$ = total number of individuals in the population with defect (1);

$D_0 = N - D_1$)

and

$$\begin{aligned} \text{Pr}[\text{having defect (2)} | \text{classified as having defect (1)}] &= \frac{N^{-1} (D_{01} p_1' + D_{11} p_1)}{\bar{p}_1} \\ &= \frac{\bar{p}_{2|1}}{\bar{p}_1}, \text{ say} \end{aligned} \quad (19.2)$$

$$\text{Hence } E[Z_2^* | Z_1] = Z_1 \bar{p}_{2|1} / \bar{p}_1 \quad (20.1)$$

and (from (7.1))

$$E[Z_2^*] = n \bar{p}_{2|1} \quad (20.2)$$

Formula (20.1) is also valid conditionally on $\underline{y} = \underline{y}$, so we have

$$\text{cov}(Z_1, Z_2^*) = \frac{\bar{p}_{2|1}}{\bar{p}_1} E[Z_1^2] - n^2 \bar{p}_{2|1} \quad (21.1)$$

and from (7.2) and (7.1)

$$\text{cov}(Z_1, Z_2^*) = \frac{n\bar{p}_2|1}{\bar{p}_1} \left\{ \bar{p}_1(1-\bar{p}_1) - \frac{n-1}{N-1} \frac{D_1}{N} \left(1 - \frac{D_1}{N}\right) (p_1 - p_1')^2 \right\} \quad (21.2)$$

We must have $\text{cov}(Z_1, Z_2^*) > 0$ because from (20.1), $E[Z_2^*|Z_1]$ is an increasing function of Z_1 .

Extension to situations in which there are $m (> 1)$ types of defects - (2), (3), ..., (m+1) - in addition to the type (1) which is inspected directly, is straightforward. The only essentially new problems are the joint distribution of Z_2^*, \dots, Z_{m+1}^* - the numbers of individuals with type (2), ..., (m+1) defects among those classified as having defect (1); and also the distributions of variables like $Z_{(ij)}^*$, the number among these individuals, with both (i) & (j) type defects ($i \neq j \geq 2$). Using an obvious notation (with subscripts 0(1) indicating absence (presence) of the corresponding type of defect) we have, for example, corresponding to (19.2)

$\text{Pr}[\text{having defects (2) and (3), but not (4), ..., (m+1)} | \text{classified as having}$

$$\text{defect (1)}] = \frac{D_{0110\dots 0} p_1' + D_{1110\dots 0} p_1}{D_{0\dots} p_1' + D_{1\dots} p_1} \quad (22)$$

where $D_{g\dots} = \sum_{a_2=0}^1 \dots \sum_{a_{m+1}=0}^1 D_g a_2 \dots a_{m+1}$ ($g = 0, 1$) ($D_{g\dots}$ is the quantity

previously represented by D_g .)

Considering, for simplicity, the case $m = 2$, we now obtain an expression for $\text{cov}(Z_2^*, Z_3^*)$. We have, analogously to (17.3)

$$Z_2^* = W_{111} + W_{011} + W_{110} + W_{010} \quad (23)$$

$$Z_3^* = W_{111} + W_{011} + W_{101} + W_{001}$$

where the W 's are independent binomial variables and the parameters of W_{ghi} are $(y_{ghi}, \delta_g p_1 + (1-\delta_g) p_1')$.

Symbolically

$$\left. \begin{matrix} Z_2^* \\ Z_3^* \end{matrix} \right\} \sim \text{Bin.}(Y_{111}, p_1) * \text{Bin.}(Y_{011}, p_1') * \left\{ \begin{matrix} \text{Bin}(Y_{110}, p_1) * \text{Bin}(Y_{010}, p_1') \\ \text{Bin}(Y_{101}, p_1) * \text{Bin}(Y_{001}, p_1') \end{matrix} \right\} \sim \Delta \text{Mult. Hypg}(n; \underline{D}, N) \quad (24)$$

(Here $\underline{y} = (Y_{111}, Y_{011}, Y_{110}, Y_{010}, Y_{101}, Y_{001})$.)

A general expression for the joint factorial moment $E[Z_2^{*(s_2)} Z_3^{*(s_3)}]$ can be obtained in the following way.

$$\begin{aligned} Z_2^{*(s_2)} Z_3^{*(s_3)} &= \left\{ \sum_{\underline{i}}^{(4)} \binom{s_2}{\underline{i}} W_{111}^{(i_1)} W_{011}^{(i_2)} W_{110}^{(i_3)} W_{010}^{(i_4)} \right\} \left\{ \sum_{\underline{j}}^{(4)} \binom{s_3}{\underline{j}} W_{111}^{(j_1)} W_{011}^{(j_2)} W_{101}^{(j_3)} W_{001}^{(j_4)} \right\} \\ &= \sum_{\underline{i}}^{(4)} \sum_{\underline{j}}^{(4)} \binom{s_2}{\underline{i}} \binom{s_3}{\underline{j}} \left\{ \sum_{g=0}^{j_1} \binom{j_1}{g} i_1^{(j_1-g)} W_{111}^{(i_1+g)} \right\} \left\{ \sum_{h=0}^{j_2} \binom{j_2}{h} i_2^{(j_2-h)} W_{011}^{(i_2+h)} \right\} \\ &\quad \times W_{110}^{(i_3)} W_{010}^{(i_4)} W_{101}^{(j_3)} W_{001}^{(j_4)} \end{aligned} \quad (25)$$

where $\sum_{\underline{i}}^{(4)} = \sum_{i_1} \sum_{i_2} \sum_{i_3} \sum_{i_4} ; \sum_{\underline{j}}^{(4)} = \sum_{j_1} \sum_{j_2} \sum_{j_3} \sum_{j_4}$ and utilizing the

$$\text{identity } a^{\binom{\alpha_1}{a}} a^{\binom{\alpha_2}{a}} = a^{\binom{\alpha_1}{a-\alpha_1+\alpha_1}} a^{\binom{\alpha_2}{a}} = \sum_{u=0}^{\alpha_2} \binom{\alpha_2}{u} a^{\binom{\alpha_1}{a}} a^{\binom{\alpha_2-u}{a}}.$$

An expression for $E[Z_2^{*(s_2)} Z_3^{*(s_3)} | \underline{y} = \underline{y}]$ is obtained by replacing $W_{ghi}^{(\alpha)}$ in (25)

by $Y_{ghi}^{(\alpha)} \tilde{p}_g^{\alpha}$ where $\tilde{p}_g = g \cdot p_1 + (1-g)p_1'$. Then taking expectations with respect to \underline{y} , we obtain

$$\begin{aligned} E[Z_2^{*(s_2)} Z_3^{*(s_3)}] &= \sum_{\underline{i}}^{(4)} \sum_{\underline{j}}^{(4)} \sum_{g=0}^{j_1} \sum_{h=0}^{j_2} \binom{s_2}{\underline{i}} \binom{s_3}{\underline{j}} \binom{j_1}{g} \binom{j_2}{h} p_1^{g+i_1+i_3+j_3} p_1^{h+i_2+i_4+j_4} \\ &\quad \times i_1^{(j_1-g)} i_2^{(j_2-h)} \cdot \frac{n^{\binom{s_2+s_3+g+h}{N}}}{N} D_{111}^{(i_1+g)} D_{011}^{(i_2+h)} D_{110}^{(i_3)} D_{010}^{(i_4)} D_{101}^{(j_3)} D_{001}^{(j_4)} \end{aligned} \quad (26)$$

If $m > 2$, D_{ghi} is replaced by $D_{ghi} \dots$; joint factorial moments of three or more Z_i^* 's ($i > 1$) can be obtained by similar techniques, though the formulas rapidly become more cumbersome.

For lower order moments, direct calculation is often simpler than using general formulas.

We now outline the calculation of $\text{cov}(Z_2^*, Z_3^*)$. We have from (23) and (24)

$$\text{cov}(Z_2^*, Z_3^* | y) = y_{111} p_1 (1-p_1) + y_{011} p_1' (1-p_1') \quad (27)$$

whence

$$E[Z_2^* Z_3^* | y] = y_{111} p_1 (1-p_1) + y_{011} p_1' (1-p_1') + \{(y_{111} + y_{110}) p_1 + (y_{011} + y_{010}) p_1'\} \\ \times \{(y_{111} + y_{101}) p_1 + (y_{011} + y_{001}) p_1'\},$$

so (using (15))

$$\text{cov}(Z_2^*, Z_3^*) = E[Z_2^* Z_3^*] - E[Z_2^*] E[Z_3^*]$$

$$= \frac{n}{N} (D_{111} p_1 (1-p_1) + D_{011} p_1' (1-p_1')) + \left\{ \frac{n}{N} - \frac{n(n-1)}{N(N-1)} \right\} (D_{111} p_1^2 + D_{011} p_1'^2) \\ + \left\{ \frac{n}{N} \frac{(n-1)}{(N-1)} - \frac{n^2}{N^2} \right\} \{(D_{111} + D_{110}) p_1 + (D_{011} + D_{010}) p_1'\} \{(D_{111} + D_{101}) p_1 + (D_{011} + D_{001}) p_1'\} \\ = \frac{n}{N} \{D_{111} p_1 (1 - \frac{n-1}{N-1} p_1) + D_{011} p_1' (1 - \frac{n-1}{N-1} p_1')\} \quad (28) \\ - \frac{n(N-n)}{N^2(N-1)} \{(D_{111} + D_{110}) p_1 + (D_{011} + D_{010}) p_1'\} \{(D_{111} + D_{101}) p_1 + (D_{011} + D_{001}) p_1'\}$$

This covariance can be positive or negative. Both Z_2^* and Z_3^* are positively correlated with Z_1 (the total number of individuals, classified as 'defective', of which Z_2^* and Z_3^* are subsets), but they may be negatively correlated in the population. The latter situation corresponds to values of D_{g10} and D_{g01} which are large relative to D_{g11} . ($g = 0, 1$).

When N is large compared with n ,

$$\text{cov}(Z_2^*, Z_3^*) \doteq n[\bar{p}_{1|11} - (\bar{p}_{1|11} + \bar{p}_{1|10})(\bar{p}_{1|11} + \bar{p}_{1|01})] \quad (29)$$

where $\bar{p}_{1|hi} = N^{-1}(D_{1hi} p_1 + D_{0hi} p_1')$.

Taking as an example $p_1 = 0.90$, $p_1' = 0.10$

$$D_{111}/N = 0.01; D_{110}/N = 0.1; D_{101}/N = 0.15$$

$$D_{011}/N = 0.01; D_{010}/N = 0.15; D_{001}/N = 0.2$$

we have

$$\bar{p}_1|111 = 0.010; \bar{p}_1|110 = 0.105; \bar{p}_1|101 = 0.155$$

so

$$\bar{p}_1|111 - (\bar{p}_1|111 + \bar{p}_1|110)(\bar{p}_1|111 + \bar{p}_1|101) < 0$$

corresponding to negative $\text{cov}(Z_2^*, Z_3^*)$.

On the other hand, if D_{111}/N and D_{011}/N are each increased to 0.05, the other parameters remaining the same, we have $\bar{p}_1|111 = 0.050$, while $\bar{p}_1|110$ and $\bar{p}_1|101$ remain unchanged, so

$$\bar{p}_1|111 - (\bar{p}_1|111 + \bar{p}_1|110)(\bar{p}_1|111 + \bar{p}_1|101) = 0.050 - (0.155 \times 0.205) > 0$$

corresponding to positive $\text{cov}(Z_2^*, Z_3^*)$.

The same formulas apply when there are $m(>2)$ types of defect other than (1), replacing D_{ghi} by $D_{ghi} \dots$.

At the cost of some elaboration in the formulas, we can allow for the possibility that presence or absence of a defect of type (2) may affect the probability of correct classification in regard to defects of type (1). Introducing the notation

${}_2p_1$ (${}_2\bar{p}_1$) for probabilities of detection of (1) in the presence (absence) of (2)

and

${}_2p_1'$ (${}_2\bar{p}_1'$) for probabilities of incorrect assignment of (1) when no (1) is present, in the presence (absence) of (2) we would still have a model of form (17) but the parameters of the binomial distributions of the

W_{gh} 's would now be $(y_{g0}, g{}_2p_1 + (1-g){}_2\bar{p}_1')$ for $h = 0; g = 1, 2$

$$(y_{g1} \cdot \delta_g \cdot \bar{z}_{p1} + (1 - \delta_g) \bar{z}_{p1}') \quad \text{for } h = 1; g = 1, 2 .$$

The probability of having defect (2), if classified as having (1) would be

$$\frac{D_{01} \cdot \bar{z}_{p1}' + D_{11} \cdot \bar{z}_{p1}}{D_{00} \cdot \bar{z}_{p1}' + D_{01} \cdot \bar{z}_{p1}' + D_{10} \cdot \bar{z}_{p1} + D_{11} \cdot \bar{z}_{p1}} = \frac{\bar{p}_{2|1}^*}{\bar{p}_1^*} \quad (30)$$

where $\bar{p}_1^* = \frac{1}{N}(D_{00} \cdot \bar{z}_{p1}' + D_{01} \cdot \bar{z}_{p1}' + D_{10} \cdot \bar{z}_{p1} + D_{11} \cdot \bar{z}_{p1}) = \text{Pr}[\text{classified as having defect (1)}]$

and $\bar{p}_{2|1}^* = \frac{1}{N}(D_{01} \cdot \bar{z}_{p1}' + D_{11} \cdot \bar{z}_{p1})$

Formulas (20.1) and (20.2) would still be valid, with \bar{p}_1 , $\bar{p}_{2|1}$ replaced by \bar{p}_1^* , $\bar{p}_{2|1}^*$ respectively.

Extension to situations with $m(>1)$ types of defect, other than the one (type (1)) which is inspected directly is, again, straightforward.

In view of the model (17) which applies, with appropriate adjustments, to all the cases mentioned above, the joint distribution of Z's and Z*'s is asymptotically multinomial as the population size N increases indefinitely, with the ratios $n:D's:N$ remaining constant, or tending to fixed values.

5. SOME APPLICATIONS

Although this paper is concerned primarily with some novel compound multivariate discrete distribution which can arise in connection with faulty inspection rather than in specific applications, we shall indicate in this section a few circumstances in which knowledge of these distributions may be useful and directly applicable to specific investigations and inquiries

The results in Section 2 are relevant to studies of robustness of multistage sample procedures to errors incurred in inspection and consequently to the

actual construction of such procedures. They would also be directly relevant to construction of tests for inequalities among the p_i 's and/or p_i' 's which could be one aspect of attempting to detect the existence of faulty items. (Evidently if p_i and/or p_i' vary with i , they cannot be identically equal to 1 or 0 respectively for all i). The distributions derived in this paper are also indirectly relevant to construction of tests of hypothesis of no faults ($p = 1, p' = 0$) assuming p_i, p_i' do not depend on i . Some attempts in this direction have been made in Johnson & Kotz (1982) while analyses of ways in which cost consideration can be allowed for in faulty inspection problems are sketched in Kotz & Johnson (1982b).

The results in Section 4 are relevant to assessment of performance of procedures for identifying individuals with defects of type (2), say (especially in those cases when these defects are not easily detectable) by observing the existence or non-existence of defects of type (1), and the robustness of this assessment to actual numbers of faults among inspected items. In these circumstances it may sometimes be appropriate to carry out a 100% inspection - that is, $n = N$ - though the more general formulas we have derived are of course of greater flexibility and are useful in various situations when total inspection is either not feasible or too costly. Indeed, studies in this direction will involve introduction of cost functions allowing for costs of sampling and losses due to erroneous retention of defective individuals of type (2) and the erroneous rejection of non-defective type (2) items. See Kotz & Johnson (1982b) for an appropriate model and some preliminary results for the case of two-stage sampling with defects of a single type. Finally questions of choice of which type(s) of defects to inspect for can also arise in this context.

6. ADDENDUM

In this paper we have supposed that inspection is on one specific type of defect, even when other types exist. Distributions arising when there is inspection for $k(\geq 2)$ types of defect will be discussed in later work. Variables arising include Z_{g_1, g_2, \dots, g_k} the total number in a random sample of size n who are judged to have g_i ($=0$ or 1) defects of type (i) , ($i=1, 2, \dots, k$), and $Z_{h_1, h_2, \dots, h_k}^*(g_1, g_2, \dots, g_k)$, the number, among these Z_{g_1, g_2, \dots, g_k} individuals who have in fact h_i ($=0$ or 1) defects of type (i) ($i = 1, 2, \dots, k$).

ACKNOWLEDGEMENTS

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

REFERENCES

- [1] Armstrong, W. (1982). A chart to describe and measure the affects of industrial errors. The Statistician, 31, 199-209.
- [2] Dorris, A.L. and Foote, B.L. (1978). Inspection errors and statistical quality control: A survey. AIIE Transactions, 10, 183-192.
- [3] Hoag, L.L., Foote, B.L. and Mount-Campbell, C. (1975). The effect of inspector accuracy on the Type I and Type II error of common sampling techniques. J. Qual. Techn., 7, 157-164.
- [4] Johnson, N.L. and Kotz, S. (1969). Distributions in Statistics - Discrete Distributions. Wiley, New York.
- [5] Johnson, N.L. and Kotz, S. (1981). "Faulty inspection distributions - some generalizations." Reliability in the Acquisition Processes (D. de Priest and R.L. Launer, Eds), pp 171-182. M. Dekker, New York.
- [6] Johnson, N.L. and Kotz, S. (1982). Detection of faulty inspection. Institute of Statistics, University of North Carolina, Mimeo Series #1505.
- [7] Kotz, S. and Johnson, N.L. (1982a). Errors in inspection and grading: Distributional aspects of screening and hierarchal screening. Commun. Statist. A11, 1997-2016.
- [8] Kotz, S. and Johnson, N.L. (1982b). Effects of false and incomplete identification of defective items on the reliability of acceptance sampling", Institute of Statistics, University of North Carolina, Mimeo Series #1386.
- [9] Rahali, B. and Foote, B.L. (1982). An approach to compensate for uncertainty in knowledge of inspection error. J. Qual. Techn. 14, 190-195.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Some Multivariate Distributions Arising in Faulty Sampling Inspection		5. TYPE OF REPORT & PERIOD COVERED Technical
7. AUTHOR(s) Norman L. Johnson & Samuel Kotz		6. PERFORMING ORG. REPORT NUMBER
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics University of North Carolina Chapel Hill, NC 27514		8. CONTRACT OR GRANT NUMBER(s) ONR Contract N00014-81-K-0301 NSF Grant MCS-8021704
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Office of Naval Research Statistics and Probability Program (Code 436)		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		12. REPORT DATE January 1983
		13. NUMBER OF PAGES
		15. SECURITY CLASS. (of this report)
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for Public Release -- Distribution Unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Binomial; Compound distributions; faulty inspection; hypergeometric; inspection sampling; multivariate distributions.		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) In the present paper we extend in two ways some results presented in Kotz and Johnson (<u>Comm. in Statistics</u> (1982), A11) relating to the study of distributional aspects of effects of errors in inspection sampling: (1) Multistage sampling with k successive samples involving the possibility of two types of errors in inspection (classifying a defective individual as non-defective, or a non-defective as defective); (2) Single-stage sampling considering several types of defects of which only one is tested on inspection. Both (1) and (2) lead to novel multivariate distribution. Their structural properties are (over)		