ON TESTS OF INDEPENDENCE IN SEVERAL DIMENSIONS

H. O. LANCASTER

(received 3 March 1959, revised 20 August 1959)

Introductory. This paper considers a canonical form, or rather a class of canonical forms, for three dimensional probability distributions subject to a rather mild restriction. These canonical forms are used to develop suitable tests of independence and lead to a consideration of the partition of χ^2 in the analysis of complex contingency tables. Where these methods and Bartlett's are both applicable it is shown that they give comparable results; but the partitioning methods are more general.

1. Historical and Introductory

Karl Pearson [25] was the first to consider the analysis of contingency tables in more than two dimensions. In [25], however, he was concerned with the identification of a certain limiting value of $\chi^2/N \equiv \phi^2$ with a determinantal function of the coefficients of correlation and did not proceed further with the interpretation of ϕ^2. In a very remarkable article, Pearson [26] returned to the contingency tables of several dimensions, which we shall refer to as complex contingency tables. He considered first of all the test of the hypothesis of complete independence, when the probability density is given by the multiplication rule. He stated that in this case χ^2 is distributed with $\alpha\beta\gamma \cdots - 1$ degrees of freedom even though he had used the sums in the rows and columns to estimate the row and column parameters and so on, so that he should really have had $\alpha\beta\gamma \cdots - (\alpha - 1) - (\beta - 1) \cdots - 1$ degrees of freedom. In his second case, he considered the cell frequencies to be given by some other law than complete independence although once again the row parameters were estimated from the data so that the sums of observed and expected were identical for rows, columns $\cdots$. As is well known he did not regard this type of restriction as causing any loss of degrees of freedom or, as he would say, of variables. In the next paragraph of [26], he considered additional restraints in a most interesting manner and shows how each restraint reduces the number of variables (that is, degrees of freedom) by unity. It is surprising that he failed to see that the same reasoning should apply to the fixing of marginal totals in a contingency table. In the third paragraph, Pearson [26] considered the frequencies in one two-dimensional face as fixed beforehand and the test is for homogeneity of distribution of a third variable; he referred to this as partial contingency. It is worthy of note that he introduced both homogeneous and non-homogeneous restraints and in the latter case reached a χ^2 which does not have the usual expectation. [26] is indeed full of important ideas. Further, although in the single variate distributions Pearson in his many papers was anxious to show that the normal curve was not the only one, yet in both [25] and [26] he is very concerned with the notion that the observations had arisen from a joint multivariate normal distribution.

Wilks [30], applying the theory of the likelihood ratio of Neyman and Pearson [22] derived a test function for a complex contingency table of $r \times s \times t$ cells, which is distributed approximately as χ^2 with $rst - r - s - t + 2$ degrees of freedom, the parameters for the rows, columns and layers being estimated from the data but the totals of the cells of the two dimensional marginal contingency tables were not taken as fixed. Bartlett [2] followed a suggestion of R. A. Fisher that he should apply the theory of [6] to the $2 \times 2 \times 2$ tables. In [6] it is assumed that the probability of a given contingency table can be factored into a chain of conditional probabilities similar to the well known factorization in a two dimensional contingency table after the manner of Yates [33] and Bartlett [3]. Bartlett [2] noted that his method would lead to computational difficulties even in the $3 \times 2 \times 2$ tables. However, Norton [23] applied the methods of [3] to an $R \times 2^k$ table.

The methods of Lombard and Doering [18] are rather outside the lines of development already mentioned. They adopted the technique of partial correlation from the normal theory and defined second order interactions in the same way as $r_{12.3}$ in the normal theory. This work has been criticised by Dyke and Patterson [5] who used a transformation to certain variables, z_i, and then carried out an analysis of variance of the z_i.

Wilson and Worcester [31] considered different models indicating differences in the mode of dependence of the three variables. They independently, as [2], introduced a cubic equation; they noted that the expectation of a cell is not exactly the same as the expression given by the solution of the cubic.

Winsor [32] re-examined the data of [18] using probit analysis, taking linear combinations of the transformed cell proportions.

Simpson [29] considered various definitions of interaction suggested by general probability theory and by the coefficients of Yule given in [12]. Lancaster [16] used the partition of χ^2 to obtain appropriate tests. Asymptotically his χ^2 was equivalent to the Bartlett χ^2 in the $2 \times 2 \times 2$ tables, [16]. He stated, unfortunately, that the cubic has only one root. An accurate statement would have been that it has only one real root in the admissible range.

Yates [34] noted that the theory of three dimensional distributions is necessary if comparisons of two treatments in a number of different trials are being compared.

Freeman and Halton [8] proposed to enumerate the exact probabilities and thereby avoid the use of χ^2. They also have used the theory of the factorisation of probabilities into chains of conditional probabilities.

Mitra [20] and Roy and Kastenbaum [27] and [28] have used the theory of Fisher [6], Barnard [1] and E. S. Pearson [24] to give generalisations of the Fisher theory to higher dimensions and distinguish between variables and categories.

We shall make further comments on the historical aspects after having set up a model for three dimensional distributions subject to a certain restriction, which enables a canonical form to be given and which is fulfilled for all distributions with a finite number of points and for very many other distributions such as the multivariate normal.

2. The probability model

In this section, we shall extend the analysis of [15] to contingency tables of higher dimensions, discussing the three-dimensional case in detail. For notational convenience we take the most general three dimensional distribution function to be $F(x, y, z)$ and marginal distribution functions to be $G(x)$, $H(y)$ and $K(z)$ and drop the variable, writing F, G, H and K. The discussion is limited to ϕ^2-bounded distributions, that is, those obeying the condition,

$$(1)\qquad \phi^2 + 1 = \int dF^2/(dGdHdK) = \int (dF/dGdHdK)^2 dGdHdK < \infty,$$

with the integral used in the sense of Hellinger as explained in [15]. The marginal distributions may have a finite or infinite number of points of increase.

Complete orthonormal sets $\{x^{(i)}\}$, $\{y^{(i)}\}$ and $\{z^{(i)}\}$ can be defined on the marginal distributions for this can be established for arbitrary statistical distributions. The product set, $\{x^{(i)}\} \times \{y^{(i)}\} \times \{z^{(i)}\}$ is then complete on the product distribution, which can be written symbolically as $G \times H \times K$. Now by (1) the ratio $\{dF/(dGdHdK)\}$ is square summable on this product distribution and so may be expanded in a series, so that

$$(2)\qquad dF = \{1 + \sum_{j,k} \rho_{0jk} y^{(j)} z^{(k)} + \sum_{i,k} \rho_{i0k} x^{(i)} z^{(k)} + \sum_{i,j} \rho_{ij0} x^{(i)} y^{(j)} + \sum_{i,j,k} \rho_{ijk} x^{(i)} y^{(j)} z^{(k)}\} dGdHdK,$$

where

$$\phi^2_{xy} + \phi^2_{xz} + \phi^2_{xy} + \phi^2_{ijk} = \phi^2, \tag{3}$$

with the ϕ's defined as in (4) below and by the Schwarz inequality each ρ is less in absolute value than unity. We do not follow the convention, possible in the two dimensional case, that the ρ's are non-negative.

Definition. The *main effects* or *zero order interactions* are defined to be the expectations of $x^{(i)}$, $y^{(i)}$ and $z^{(i)}$ respectively. (In the theoretical distributions these are all zero). The *first order interactions* are defined here to be the expectation of the forms, $x^{(i)}y^{(j)}$, $x^{(i)}z^{(k)}$ and $y^{(j)}z^{(k)}$, the indices each being greater than zero. These expectations are ρ_{ij0}, ρ_{i0k} and ρ_{0jk} respectively. Similarly the *second order interactions* are defined as the expectations of the forms, $x^{(i)}y^{(j)}z^{(k)}$, namely ρ_{ijk}. The process of definition can be extended into higher dimensions. We further define,

$$\phi^2_{xy} = \sum \rho^2_{ij0}, \quad \phi^2_{yz} = \sum \rho^2_{0jk}, \quad \phi^2_{xz} = \sum \rho^2_{i0k} \tag{4}$$

with summation over the non-zero indices. We prove later that these quantities are invariants under certain transformations. In the $2 \times 2 \times 2$ table we write ϕ_{xy} in place of ρ_{110}.

$$\phi^2_{xyz} = \sum_{i,j,k} \rho^2_{ijk}, \tag{5}$$

is similarly defined for the three dimensions. The process may be continued if required into higher dimensions. The estimates of the ρ's will be the mean value of the forms in the sample e.g. $Sx^{(i)}y^{(j)}/N$.

Theorem 1. *If F is a ϕ^2-bounded distribution function, then dF can be expanded in the form* (2). *The sum of squares of the coefficients is given by the Parseval equality,*

$$\phi^2 \equiv \phi^2_{xy} + \phi^2_{yz} + \phi^2_{xz} + \phi^2_{xyz}. \tag{6}$$

Proof. This follows from the general theory of orthonormal sets.

Theorem 2. *A necessary and sufficient condition for complete independence of the three marginal variables is that ϕ^2 should be zero.*

Proof. This is obvious but the theorem is stated because it is the basis for some tests of independence.

Theorem 3. *For the class of three dimensional ϕ^2-bounded distributions, the two dimensional marginal distributions are ϕ^2-bounded and the bivariate element of distribution can be expanded in the form,*

$$dL(x, y) = \{1 + \sum_{i,j} \rho_{ij0} x^{(i)} y^{(j)}\} \, dGdH, \tag{7}$$

almost everywhere.

Proof. This follows from (2) by integrating out the variable, z.

THEOREM 4. *For any choice of complete sets of orthonormal functions on the marginal distributions, the following are invariants,* ϕ^2_{xy}, ϕ^2_{yz}, ϕ^2_{xz}, ϕ^2_{xyz} *and consequently also* ϕ^2.

PROOF. Suppose that a bi-unique transformation is made of the form, $\{x^{(i)}\} \to \{x^{*(i)}\}$, for definiteness, on one of the marginal distributions. Let us consider the coefficients of $x^{(i)}y^{(j)}$ for a fixed value of j. In the first system, the multipiier of $y^{(j)}$ is

$$l(x, j) = \sum_i \rho_{ij0} x^{(i)} = \rho^T x, \quad \text{with} \quad \sum_i \rho^2_{ij0} \leqq \phi^2_{xy} < \infty. \tag{8}$$

Now let this function, $l(x, j)$ be written in the new set of variables by means of the orthogonal transformation,

$$\boldsymbol{x} = \boldsymbol{R}\boldsymbol{x}^*, \tag{9}$$

$\boldsymbol{R}$ orthogonal. Then the new coefficients will be given by the elements of $\rho^T\boldsymbol{R} \equiv \rho^{*T}$. But since $\boldsymbol{R}$ is orthogonal, the sums of squares of the ρ's, namely $\sum_i \rho^2_{ij0} = \sum_i \rho^{*2}_{ij0}$ will be invariant for a fixed j. Consequently ϕ^2_{xy} is invariant.

Note that the constant term of either orthogonal sets is not involved. By varying x, y and z in turn we can prove the whole theorem.

Let us consider the various possibilities of the sums ϕ^2, ϕ^2_{xy}, ϕ^2_{xyz}, having zero or non-zero values. Instead of writing ϕ_{xyz} we shall write (xyz) in the following. The possibilities are that for some x, y, z, each non-zeıo,

(i) $(xyz) = 0, \quad (xy0) = (0yz) = (x0z) = 0$
(ii) $(xyz) = 0, \quad (xy0) \neq 0, \quad (0yz) = (x0z) = 0$
(iii) $(xyz) = 0, \quad (xy0) \neq 0, \quad (0yz) \neq 0, \quad (x0z) = 0$
(iv) $(xyz) = 0, \quad (xy0) \neq 0, \quad (0yz) \neq 0, \quad (x0z) \neq 0.$

(Two more possibilities are open for each of (ii) and (iii) by a permutation of the variables x, y and z).

(v) $(xyz) \neq 0, \quad (xy0) = (0yz) = (x0z) = 0$
(vi) $(xyz) \neq 0, \quad (xy0) \neq 0, \quad (0yz) = (x0z) = 0$
(vii) $(xyz) \neq 0, \quad (xy0) \neq 0, \quad (0yz) \neq 0, \quad (x0z) = 0$
(viii) $(xyz) \neq 0, \quad (xy0) \neq 0, \quad (0yz) \neq 0, \quad (x0z) \neq 0.$

(Two more possibilities are open for each of (vi) and (vii) by a permutation of the variables x, y and z).

EXAMPLE (i) There exist three dimensional distributions corresponding to each of the possibilities (i) to (viii) listed above. We take each distribution to consist of equal masses at the points ± 1 and define complete orthonormal sets for each variable of the form, $x^{(0)} = 1$, $x^{(1)} = x$, then

$$p(x, y, z) = \tfrac{1}{8}(1 + \rho_{110}xy + \rho_{011}yz + \rho_{101}xz + \rho_{111}xyz) \tag{10}$$

is a frequency distribution in three dimensions if each $|\rho| \leqq 1$, and

https://doi.org/10.1017/S1446788700025593 Published online by Cambridge University Press

(11) $$|\rho_{110}| + |\rho_{101}| + |\rho_{011}| + |\rho_{111}| \leqq 1$$

under these conditions no frequency element is negative.

TABLE 1

Independence table for a three dimensional statistical distribution

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Interactions				Values of the ϕ^2				
ϕ^2_{xyz}	0*	0	0	0	+**	+	+	+
ϕ^2_{xy}	0	+	+	+	0	+	+	+
ϕ^2_{yz}	0	0	+	+	0	0	+	+
ϕ^2_{xz}	0	0	0	+	0	0	0	+
Variables				Dependence Values				
x, y	I	D	D	D	I	D	D	D
y, z	I	I	D	D	I	I	D	D
x, z	I	I	I	D	I	I	I	D
$x, (y, z)$***	I	D	D	D	D	D	D	D
$y, (x, z)$	I	D	D	D	D	D	D	D
$z, (x, y)$	I	I	D	D	D	D	D	D

* 0 signifies $\phi^2_{xyz} = 0$ *et cetera*.

** + signifies $\phi^2_{xyz} \neq 0$ *et cetera*

*** I in this part of the table is to be read, x is independent of the joint distribution of y and z. D signifies that this is not so.

The different cases can now be interpreted with the aid of a table in which we write I for independence and D for dependence. We shall say that a dependency exists if the independence conditions are not fulfilled. The various cases given above may be considered in turn:

(i) Complete independence.

(ii) z is independent separately of x and of y and of them jointly. All other possible relations show dependencies.

(iii) x is independent of z. All other relations are dependencies.

(iv) No pair of marginal variables is mutually independent. No variable is independent of the joint distribution of the other two.

In (v) to (viii) no variable is independent of the joint distribution of the other two.

(v) The marginal variables are independent in pairs but no variable is independent of the other two jointly.

(vi) Two pairs of marginal variables are mutually independent but no variable is independent of the other two jointly.

(vii) One pair of marginal variables are mutually independent.

(viii) All relations are dependencies.

The relations may be gathered together as a theorem.

THEOREM 5. *ϕ^2-bounded three dimensional distributions can be classed into 8 types according to whether the ϕ^2_{xyz} is zero or not and according to the number of the ϕ^2_{xy}, ϕ^2_{yz} and ϕ^2_{xz} which are zero.*

The dependence or independence can be read off from Table 1.

Arbitrary three-dimensional distributions can be set up as in section 7 of (15).

EXAMPLE. (ii). The multivariate normal distribution is often considered as being relevant, [25], [26] and [18]. Expansions can be given as in [11] of the frequency function in the form (2). Unfortunately these expressions have not the simple form of the Mehler expansion (the tetrachoric in other terminology). It is easily seen that the partial correlation technique may be quite irrelevant as the ρ_{ijk} may have no necessary relation with the ρ_{ij0} *et cetera* if the distribution is not jointly normal. In the case of a joint normal distribution,

$$\phi^2 = |(\mathbf{1} + \mathbf{P})(\mathbf{1} - \mathbf{P})|^{-\frac{1}{2}} - 1, \quad \phi^2_{xy} = r^2_{xy}(1 - r^2_{xy})^{-1}, \tag{12}$$

with similar expressions for ϕ^2_{xz} and ϕ^2_{yz}. ϕ^2_{xyz} can be obtained by subtraction.

In (12) the elements of $\mathbf{P}$ are given by $p_{ii} = 0$, $p_{ij} = r_{ij}$ for $i \neq j$, $\mathbf{1} + \mathbf{P}$ and $\mathbf{1} - \mathbf{P}$ being positive definite.

3. The Theory underlying Bartlett's Method

Interesting theorems, due substantially to Fisher [6] and Wilson and Worcester [31], which we cite without proof, treat the conditional distributions arising in the study of contingency tables.

THEOREM 6. *In the fourfold (i.e.* 2×2) *table, a necessary and sufficient condition that the distribution of the cell totals, given the marginal totals, should be independent of the value of the parameters of the marginal distribution is that*

$$p_1 p_4 = p_2 p_3 \tag{13}$$

THEOREM 7. *In the* $2 \times 2 \times 2$ *table, a necessary and sufficient condition that the distribution of the cell totals, given the marginal totals, should be independent of the parameters of the marginal distribution is that*

$$p_1 p_4 p_6 p_7 = p_2 p_3 p_5 p_8 \tag{14}$$

These results can be extended to $r \times s$ and $r \times s \times t$ tables respectively, by constructing all possible 2×2 and $2 \times 2 \times 2$ tables from the given marginal totals.

This theory of the probability, conditional on the marginal totals, is the basis of Bartlett's treatment of the $2 \times 2 \times 2$ tables. Let us apply his method to a $2 \times 2 \times 2$ table supposed to be obtained from a sample of size of arbitrarily large, N. In doing so we are really studying the theoretical distribution. Bartlett's method leads us to add a quantity to the cell number in one cell and subtract it from others. In the theoretical case such a constant is of the form,

$$\text{a constant} \times x^{(1)}y^{(1)}z^{(1)}dGdHdK,$$

which is of sufficient importance to write in as a theorem.

THEOREM 8. *Let $U(x, y, z)$ be a function which is square summable with respect to the product of the marginal distributions and such that the expression, $U(x, y, z)\ dG(x)$ is zero for almost all values of (y, z) and a similar relation holds for the other two variables. Then*

$$U(x, y, z) = \sum_{i,j,k} \tau_{ijk} x^{(i)} y^{(j)} z^{(k)}, \tag{15}$$

summation being over all positive integers i, j, k.

PROOF. By the same type of analysis as justifies (2), we minimise the integral of the square of U less a (possibly infinite) series of orthonormal functions. We find that the restrictions introduced ensure that the value of the integrals of $Ux^{(i)}y^{(j)}$ and like terms vanish.

$$\tau_{ijk} = \int U(x, y, z)\, x^{(i)} y^{(j)} z^{(k)} dGdHdK. \tag{16}$$

Theorem 8 is true both for theoretical distributions and for empirical distributions, where marginal parameters have been estimated from the data.

COROLLARY 1. If the sample size N of a $2 \times 2 \times 2$ table is imagined to be indefinitely large, Bartlett's χ^2 is (asymptotically) $N\phi^2_{xyz}$.

PROOF. We have only two points with positive weights on each of the marginal distributions. We have then only one orthonormal function possible; we can take it without loss of generality to take this function as the marginal variable. The theoretical distribution can then be written,

$$dF = (1 + \phi_{xy}xy + \phi_{xz}xz + \phi_{yz}yz + \phi_{xyz}xyz)dGdHdK. \tag{17}$$

As N is indefinitely large, the cell contents all approximate to the theoretical. The cell contents are approximately $NdF(x, y, z)$.

Theorem 8 enables us to write the quantity to be added and subtracted to the cells as $N\Delta xyz\, dGdHdK$, where Δ has to be found by the solution of a certain cubic. After cancellations of powers of N and terms of the type $dG(x)$, we get the difference of two products, which we write

$$\Big\{ \prod_{1,4,6,7} - \prod_{2,3,5,8} \Big\} \{1 + \phi_{xy}xy + \phi_{xz}xz + \phi_{yz}yz + (\phi_{xyz} \pm \Delta)xyz\} = 0 \tag{18}$$

where the first product is taken over cells numbers 1, 4, 6, 7 and the second over 2, 3, 5 and 8. If the first order interactions ϕ_{xy}, ϕ_{xz} and ϕ_{yz} are zero, this has an exact solution

$$\Delta = \phi_{xyz}. \tag{19}$$

If they are not quite large (19) is approximately true. If two of the first order ϕ's are zero (19) is exact once again. By Bartlett's rule for computing the χ^2, we have

$$\begin{aligned}\chi^2 &= N\int (\Delta xyzdGdHdK)^2/\{(1+\phi_{xy}xy+\phi_{xz}xz+\phi_{yz}yz)dGdHdK\} \\ &= N\int \Delta^2x^2y^2z^2/\{1+\phi_{xy}xy+\phi_{xz}xz+\phi_{yz}yz\}\, dGdHdK \\ &\simeq N\Delta^2 \simeq N\phi^2_{xyz}.\end{aligned} \tag{20}$$

The denominator may be taken as approximately unity. Moreover, the fact that these denominators vary about unity will make this approximation more realistic. In fact, the ratio of the approximate value to the true value may be expected to differ from unity by a quantity of the order of the square of the largest of the first order ϕ's. But $N\phi^2_{xyz}$ is the second order χ^2 obtained by partition in the manner of [16].

Corollary 2. For finite sample size, N, of a $2 \times 2 \times 2$ table, Bartlett's χ^2 approximates to $N\phi^2_{xyz}$ if the theoretical marginal parameters are available or to Nr^2_{xyz} if the parameters are estimated and so to the χ^2 of the second order interaction of [16]. Or better, Bartlett's χ may be regarded as a non-central normal variable with variance not greatly different from unity.

Proof. The proof goes through as in Corollary 1; the distribution can be reduced to canonical form (17) with a multiplier, N. If Δ is chosen consistent with the rules laid down by Bartlett, namely solving (18), then the solution is $\Delta = r_{xyz}$, where r_{xyz} is the observed value of the second order interaction with the marginal parameters estimated from the data. But

$$N^{1/2} r_{xyz} = N^{-1/2} \mathop{S}_{1}^{N} x_j y_j z_j, \tag{21}$$

which can be shown to be the second order χ with parameters estimated from the data. This identification follows essentially from the observation that the second row of the orthogonal matrices used to partition χ^2 in [16] can be written $x\sqrt{dG(x)}$ in the notation of this paper.

4. A Random Sampling Experiment

Random sampling numbers were used to imitate sampling from three dimensional distributions. For computational convenience, the distributions were specialised so that the marginal frequencies were non-zero at only two points, each associated with a probability, $\frac{1}{2}$. There is only one orthonormal function on such a distribution and without loss of generality we take it to be the marginal variable, x, y or z, as the case may be. To obey the orthonormal conditions these variables can take only the values ± 1. In the first two sets of drawings the sample total, N, was taken to be a Poisson variable

https://doi.org/10.1017/S1446788700025593 Published online by Cambridge University Press

with parameter 640; in the third set the parameter has the value 720. The first set of 10 drawings were made to test whether large first order interactions would be likely to cause spuriously high values of the second order interactions. The theoretical values of $\frac{1}{4}$ were given to each of the first order correlations and zero to the second order correlation. In the second set, $\rho_{110} = \rho_{011} = \frac{1}{4} = \rho_{111}$, other ρ's all zero. In the third set $\rho_{111} = \frac{1}{3}$, all other ρ's being zero. The usual χ^2 of a contingency table were computed using first

TABLE 2

A sampling experiment

Drawing	First Series			Second Series			Third Series		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
1	+3.67	+3.07	+3.50	−0.23	−0.60	−0.31	+3.06	+3.06	+3.09
2	−0.12	−0.26	−0.24	+0.28	+0.11	+0.45	+0.30	+0.34	+0.36
3	+1.65	+1.40	+1.65	+0.22	+0.39	+0.53	−1.74	−1.75	−1.75
4	−0.52	−0.39	−0.45	+0.04	−0.27	+0.10	−0.95	−0.90	−0.89
5	+1.40	+1.40	+1.50	−0.66	−0.96	−0.81	+0.69	+0.40	+0.61
6	+0.44	+0.38	+0.44	+1.53	+1.64	+1.82	−1.52	−1.25	−1.53
7	−0.75	−1.01	−0.99	−0.17	−0.29	−0.06	−0.88	−0.85	−0.85
8	+0.76	+0.04	+0.26	−1.10	−1.11	−0.97	+0.84	+0.81	+0.86
9	−0.44	−0.51	−0.52	+0.29	−0.35	0.00	+0.76	+0.80	+0.74
10	−1.17	−1.41	−1.39	−0.21	−0.32	−0.11	−1.49	−1.56	−1.51

First Series $\rho_{110} = \rho_{101} = \rho_{011} = 0.25$, $\rho_{111} = 0$
Second Series $\rho_{110} = 0.25 = \rho_{111}$
Third Series $\rho_{111} = \frac{1}{3}$. All other ρ's zero.
(1) Second order interaction χ less $\rho_{111} N$, theoretical parameters.
(2) Second order interaction χ less $\rho_{111} N$, estimated parameters.
(3) Second order interaction χ less $\rho_{111} N$, by Bartlett's method.

theoretical parameters $p_{1..} = \frac{1}{2}$, *et cetera* and was also shown partitioned into the seven components [16]. The marginal parameters were estimated from the data and a second set of four components computed for the first and second order interactions [16]. Finally Bartlett's χ^2 was computed. In the first set, there was only one value of the second order χ^2 above three, where the theoretical parameters gave a value 13.48, the estimated parameters 9.42 and Bartlett's χ^2 was 12.25. The χ^2 of the first order interaction in every case in this set was greater than 29.00. There was good agreement between the two second order interaction χ^2 and Bartlett's χ^2 in practically every case. We prove the proposition in our next section, that the square root of the χ^2 with an appropriate convention for the sign is approximately distributed as a non-central normal variable, with unit variance. We, therefore, give the comparisons of the form,

$$\zeta = \sqrt{\chi^2} - \rho_{111}\sqrt{N}, \tag{22}$$

in Table 2. The values given there seem quite compatible with the hypoth-

https://doi.org/10.1017/S1446788700025593 Published online by Cambridge University Press

esis that, for either of the three test functions ζ can be regarded as a standardised normal deviate. The sign of $\sqrt{\chi^2}$ is to be taken as the sign of the difference, observed minus expected, in the first cell.

5. On the Solution of Bartlett's Cubic

In [16], I considered the solution of Bartlett's cubic with a null hypothesis of complete independence and showed that the approximations, discussed here in Corollaries 1 and 2 to Theorem 8, would hold but stated that the cubic had only one real root. The correct statement is that there can only be one admissible real solution since any other would give negative values to some of the expected frequencies. Let us consider the solution of

$$(23) \qquad \begin{aligned} 0 = Q &\equiv (a_1+\delta)(a_4+\delta)(a_6+\delta)(a_7+\delta) - (a_2-\delta)(a_3-\delta)(a_5-\delta)(a_8-\delta) \\ &= (\alpha_1+\delta)(\alpha_2+\delta)(\alpha_3+\delta)(\alpha_4+\delta) - (\beta_1-\delta)(\beta_2-\delta)(\beta_3-\delta)(\beta_4-\delta) \\ &= A-B \end{aligned}$$

where the α's and β's are the a's arranged in descending order. Now for $\delta = -\alpha_4$, $A = 0$; and A increases steadily zero as δ increases from $-\alpha_4$. For δ increasing from $-\alpha_4$ to $+\beta_4$, B is positive but decreasing since each term is positive but decreasing. Q is therefore monotonically increasing for δ in the range, $-\alpha_4$ to $+\beta_4$ from a negative value to a positive value. It has exactly one real root in this range. No other value of δ gives admissible values to the expected numbers in the cells. For $\delta < -\alpha_4$, the cell corresponding to α_4 would have a negative expectation. For $\delta > \beta_4$, the cell corresponding to β_4 would have a negative expectation. It is easy to find values for the α's and β's, which give such inadmissible values to δ as well as the admissible one. The form (23) is the most convenient for numerical solution, since trial integral values are given to δ and then better approximations are obtained by linear interpolation and iteration.

6. Computation of the Component χ^2 of a $2\times2\times2$ Table

To avoid subscripts in this section we consider the two layers of a $2 \times 2 \times 2$ table and write them alongside one another

$$\begin{matrix} a & b & \qquad & e & f & \qquad & r \\ c & d & \qquad & g & h & \qquad & R. \end{matrix}$$

We take r as the total of the first rows, R as the total of the second, $r + R = N$. For columns we similarly take s and S and for layers t and T.

The total χ^2 is obtained by computing first N^2/r and N^2/R and using these factors to multiply the squares of the elements a, be and f and c, d, g and h respectively. The items in each column are added and the sums divided by factors ts, tS, Ts and TS respectively. The sum of these four quotients is

https://doi.org/10.1017/S1446788700025593 Published online by Cambridge University Press

$\chi^2 + N$. For the interaction χ^2, we multiply the rows by R and r and form the differences of the form, $Ra - rc$.

The second order χ is then $(Ra - rc)\ TS - (Rb - rd)Ts - (Re - rg)tS + (Rf - rh)ts$ divided by $\sqrt{(NRrSsTt)}$.

The products, Rr, Ss and Tt, will have been computed in any case to obtain the χ^2 of the marginal tables. A check on the computations is now possible. This method is one suggested by equation (12) of [10]. A generalisation of this method to complex contingency of more than eight cells is possible but it is sufficient to compute the overall χ^2, the χ^2 of the marginal contingency tables and then obtain the second order interaction χ^2 by subtraction. In this latter method there is no automatic check on the computations.

7. Discussion

In the test of discrete distributions one has the choice of combinatorial and approximating tests such as χ^2. We shall not consider the combinatorial methods here and refer the reader to [1], [3], [8], [20], [21], [24], [27], [28], [29], [31], which cover the ground and give further references. Cochran [4] gives a method of treating a binomial or Poisson variable defined over a two-dimensional grid, which can be regarded as closely related to the methods suggested by our methods here. Similarly Dyke and Patterson [5] used a transformation, $z = \frac{1}{2}\log_e\{p/q\}$, which gives in effect the same result in a complex table of dimensions $r \times s \times 2$, since the binomial variable and z when normalised must give the same orthonormal function. A similar remark will apply to the probit analysis of Winsor [32]. The methods developed here enable us to give precise tests of significance for each of the hypotheses, $\rho_{ijk} = 0$, where no two of the i, j, k are zero, which are (asymptotically) independent of one another. The testing of all second order interactions has been considered by Fog [7] and Mood [21], but they have not given a method of obtaining individual degrees of freedom nor linked the test up with a canonical form of distribution. Wilks [30] has given a likelihood test which combines the first and second order interaction χ^2. It may be said that where the papers [5], [7], [21], [30] and [32] deal with the same test, they are in agreement with the methods of this paper and the chief differences are in the amount of differentiation of the hypotheses to be tested. Garner and McGill [9] and McGill [19] deal with the test of interactions from the viewpoint of information theory and the likelihood ratio test.

The present paper may be considered to be a continuation of the ideas of Irwin [10] and my earlier papers [13] to [16], in which explicit orthogonal transformations are given to derive expressions for individual degrees of freedom. It is pointed out in [15] that a passage from orthogonal matrices to orthogonal functions on a finite number of points is easy; in fact, the functions are elements of a row of an orthogonal matrix divided by the corresponding element of the first row. The generalisation from one and two dimensions to higher is carried out by the use of direct products of matrices. To consider the transition from finite numbers of points to continuous distributions, we have used orthonormal functions and this has led to the introduction of the notion of the generalised correlation coefficients, $\rho_{ijk\ldots}$. To obtain the higher order interactions and so to test whether the generalised correlation coefficients, $\rho_{ijk\ldots}$ are zero or not, we have introduced the direct product of sets of orthonormal functions of the form $\{x^{(i)}\} \times \{y^{(j)}\} \times \{z^{(k)}\}$ which enables us to test all possible departures from independence in a very wide class of distributions, those which are termed ϕ^2-bounded. The development of these ideas in detail will be given in a later paper [17]. It may be mentioned here that in the non-null case, the χ^2-variables obtained will have a non-central distribution and the $N^{1/2}$ times square roots of the ϕ^2_{xy}, ϕ^2_{xz} and ϕ^2_{yz}, defined in (2) above, will be the parameters of non-centrality. Similarly $N\phi^2_{xyz}$ will be the parameter for the χ^2 of the second order interactions and so on for higher dimensions. $N\phi^2$, the sum of all the component $N\ \phi^2$'s, will be the parameter of the non-central total χ^2 in the non-null case. The likelihood ratio when null and non-null cases are being compared will be given by the series in (2) above. So that the present analysis leads to easy identifications of χ^2 with the likelihood tests of [22] and [30].

References

[1] Barnard, G. A., Significance tests for 2×2 tables, Biometrika 34, (1947) 123—138.
[2] Bartlett, M. S., Contingency table interactions, J. roy. statist. Soc. Suppl. 2 (1935), 248—252.
[3] Bartlett, M. S., Properties of sufficiency and statistical tests, Proc. Roy. Soc. A, 160 (1936), 268—282.
[4] Cochran, W. G., The analysis of variance when experimental errors follow the Poisson or binomial laws, Ann. math. Statist. 11 (1940), 335—347.
[5] Dyke, G. V. and Patterson, H. D., Analysis of factorial arrangements when the data are Proportions, Biometrics, 8 (1952), 1—12.
[6] Fisher, R. A., The logic of inductive inference, J. roy. statist. Soc., 98 (1935), 39—54.
[7] Fog, D., Contingency tables and approximate χ^2 distributions, Math. Scand. 1 (1953), 93—103.
[8] Freeman, G. H. and Halton, J. H., Note on the exact treatment of contingency, goodness of fit, and other problems of significance, Biometrika, 38 (1951), 141—149.
[9] Garner, W. R. and McGill, W. J., The relation between information and variance analyses. Psychometrika, 21 (1956), 219—228.
[10] Irwin, J. O., A note on the subdivision of χ^2 into components, Biometrika, 36 (1949), 130—134.
[11] Kendall, M. G., Proof of relations connected with the tetrachoric series and its generalization, Biometrika, 32 (1941), 196—198.
[12] Kendall, M. G., The Advanced Theory of Statistics, Vol. 1, C. Griffin & Co., London (1943).
[13] Lancaster, H. O., The derivation and partition of χ^2 in certain discrete distributions, Biometrika 36 (1949), 117—129.

[14] Lancaster, H. O., Some properties of the bivariate normal distribution considered in the form of a contingency table. Biometrika, 44 (1957) 289—292.

[15] Lancaster, H. O., The structure of bivariate distributions. Ann. math. Statist. 29 (1958), 719—736.

[16] Lancaster, H. O., Complex contingency tables treated by the partition of χ^2, J. roy. statist. Soc., Ser. B, 13 (1951), 242—249.

[17] Lancaster, H. O., (unpublished), Analogues of the Pearson χ^2.

[18] Lombard, H. L. and Doering, C. R., Treatment of the fourfold table by partial association and partial correlation as it relates to public health problems, Biometrics, 3 (1947), 123—128.

[19] McGill, W. J., Multivariate information transmission, Psychometrika, 19 (1954), 97—116.

[20] Mitra, S. K., Contributions to the statistical analysis of categorical data, Instit. Statistics, Univ. Nth. Carolina (1955), Mimeo series 142.

[21] Mood, A. M., Introduction to the Theory of Statistics, McGraw-Hill Book Co. Inc., New York (1950).

[22] Neyman, J. and Pearson, E. S., On the use and interpretation of certain test criteria, Part II, Biometrika. 20 A (1928) 263—294.

[23] Norton, H. W., Calculation of chi-square for complex contingency tables, J. Amer. statist. Assoc. 40 (1945), 251—258.

[24] Pearson, E. S., The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table, Biometrika, 34 (1947), 139—167.

[25] Pearson, K., Mathematical contributions to the theory of evolution — XIII. On the theory of contingency and its relation to association and normal correlation. Drapers' Co. Res. Memoirs (1904) Biometric Series 1, pp 34.

[26] Pearson, K., On the general theory of multiple contingency with special reference to partial contingency, Biometrika, 11 (1916), 145—158.

[27] Roy, S. N. and Kastenbaum, M. A., On the hypothesis of 'no interaction' in a multi-way contingency table, Ann. math. Statistics, 27 (1956), 749—756.

[28] Roy, S. N. and Kastenbaum, M. A., A generalization of analysis of variance and multivariate analysis to data based on frequencies in qualitative categories or class intervals, Instit. Statistics, Univ. Nth. Carolina mimeo. No. 131.

[29] Simpson, E. H., The interpretation of interaction in contingency tables, J. roy. statist. Soc., Ser. B. 13 (1951), 238—241.

[30] Wilks, S. S., The likelihood test of independence in contingency tables, Ann. Math. Statistics, 6 (1935), 190—196.

[31] Wilson, E. B. and Worcester, J., The association of three attributes, Proc. nat. Acad. Sci. U.S.A. 28 (1942), 384—390.

[32] Winsor, C. P., Factorial analysis of a multiple dichotomy, Hum. Biol., 20 (1948), 195—204.

[33] Yates, F., Contingency tables involving small numbers and the χ^2 test, J. roy. statist. Soc. Suppl. 1 (1934), 217—235.

[34] Yates, F., A note on the application of the combination of probabilities test to a set of 2×2 tables, Biometrika, 42, (1955), 404—411.

Department of Mathematical Statistics
University of Sydney.