EXACT SIGNIFICANCE TESTS FOR CONTINGENCY TABLES EMBEDDED IN A 2^n CLASSIFICATION

MARVIN ZELEN
STATE UNIVERSITY OF NEW YORK AT BUFFALO

1. Introduction

This paper considers the analysis of multidimensional contingency tables when the contingency table can be regarded as being embedded in a 2^n factorial classification. The model assumes that the response variable is binary and is observed over n factors each at two levels. This gives rise to 2^{n-1} 2×2 contingency tables. The theoretical development is in the spirit of the Fisher-Irwin treatment of the 2×2 table. The work reported here can be regarded as a generalization and extension of their work.

The new techniques for analyzing contingency tables derived here are based on conditional reference sets. This allows derivation of exact tests of significance for testing interactions arising in a contingency table context. These tests are conditional tests and have the property that they are uniformly most powerful unbiased tests.

Although this paper only discusses binary response random variables embedded in a 2^n classification, the methods are readily extended to multinomial response embedded in an arbitrary cross classification structure. In a later paper, analyses for more general contingency tables will be developed.

The classical method for analyzing the interactions associated with a complex classification is based on chi square goodness of fit tests. More recently Kullback and his associates [5], [6] have used the ideas of information theory to analyze multidimensional contingency tables. These techniques are equivalent to likelihood ratio tests. However, both the chi square and likelihood ratio techniques are based on asymptotic distributions. The methods of analysis which use a logit model or a multiplicative model for the probability of a response also are based on asymptotic theory. It is interesting that recent reviews of the analysis of contingency tables do not refer to any exact tests for testing interactions (see Lewis [10], Goodman [4], and Plackett [11]).

This work was supported by Public Health Service Research Grant No. CA-10810 from the National Cancer Institute.

2. Preliminary results

The analysis of contingency tables will be made using the analogue of linear regression for the logistic model. These results have served as the basis of several papers by Cox [2], [3] dealing with the analysis of quantal response data and are implicit in his work. In this section the results are summarized for completeness. Let $\{Y_i\}$ be a sequence of independent random variables such that $\theta_i = P\{Y_i = 1\}$ and $1 - \theta_i = P\{Y_i = 0\}$. Let $\mathbf{x}_i' = (x_{i,1}, x_{i,2}, \cdots, x_{i,p})$ be a vector of known constants and $\boldsymbol{\beta}' = (\beta_1, \beta_2, \cdots, \beta_p)$ be a vector of unknown parameters. The θ_i will be assumed to have the form

$$\theta_i = \frac{\exp\{\boldsymbol{\beta}'\mathbf{x}_i\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{x}_i\}} \qquad \text{for} \quad i = 1, 2, \cdots, n. \tag{2.1}$$

If λ_i is defined by $\lambda_i = \log\{\theta_i/(1 - \theta_i)\}$, then we have the model

$$\lambda_i = \boldsymbol{\beta}'\mathbf{x}_i = \sum_{j=1}^{p} \beta_j x_{i,j} \qquad \text{for} \quad i = 1, 2, \cdots, n. \tag{2.2}$$

The joint frequency function of $\{Y_i\}$ is

$$f(y_1, y_2, \cdots, y_n) = \prod_{i=1}^{n} \theta_i^{y_i}(1 - \theta_i)^{1-y_i} = \frac{\exp \sum_{j=1}^{p} \beta_j t^{(j)}}{\prod_{i=1}^{n} (1 + \exp\{\boldsymbol{\beta}'\mathbf{x}_i\})} \tag{2.3}$$

where $t^{(j)} = \Sigma_{i=1}^{n} x_{i,j} y_i$. Hence $\mathbf{t}' = (t^{(1)}, t^{(2)}, \cdots, t^{(p)})$ are jointly sufficient statistics for $\boldsymbol{\beta}' = (\beta_1, \beta_2, \cdots, \beta_p)$. The frequency function of the sufficient statistics is

$$f(\mathbf{t}) = P\{T^{(1)} = t^{(1)}, T^{(2)} = t^{(2)}, \cdots, T^{(p)} = t^{(p)}\} = \frac{C(\mathbf{t}) \exp \sum_{j=1}^{p} t^{(j)}\beta_j}{\prod_{i=1}^{n} (1 + \exp\{\mathbf{x}_i'\boldsymbol{\beta}\})} \tag{2.4}$$

where $C(\mathbf{t}) = C(t^{(1)}, t^{(2)}, \cdots, t^{(p)})$ is the number of ways of permuting $y_1, y_2, \cdots, y_n$ such that $T^{(1)} = t^{(1)}, T^{(2)} = t^{(2)}, \cdots, T^{(p)} = t^{(p)}$. The combinatorial coefficient $C(\mathbf{t})$ may be found as the coefficient of

$$\zeta_1^{t^{(1)}} \zeta_2^{t^{(2)}} \cdots \zeta_p^{t^{(p)}} \tag{2.5}$$

in the generating function

$$\varphi(\zeta) = \prod_{i=1}^{n} \left(1 + \zeta_1^{x_{i,1}} \zeta_2^{x_{i,2}} \cdots \zeta_p^{x_{i,p}}\right). \tag{2.6}$$

If one is interested in making an inference about only a single β_i, say β_p, one can use the distribution of $T^{(p)}$ conditional on $T^{(j)} = t^{(j)}$ for $j = 1, 2, \cdots, p - 1$. This results in

$$f\left(t^{(p)} \middle| t^{(1)}, \cdots, t^{(p-1)}\right) = \frac{C\left(t^{(1)}, t^{(2)}, \cdots, t^{(p)}\right) \exp\{t^{(p)}\beta_p\}}{\sum_z C\left(t^{(1)}, t^{(2)}, \cdots, t^{(p-1)}, z\right) \exp\{z\beta_p\}}, \tag{2.7}$$

where the summation in the denominator is taken over the range of $t^{(p)}$. Note that (2.4) is of the exponential form. Hence use of the conditional distribution given by (2.7) results in uniformly most powerful unbiased tests for testing the null hypothesis $H_0: \beta_p = 0$ versus one sided or two sided alternatives (see Lehmann [9]). Under $H_0: \beta_p = 0$, the conditional distribution of $T^{(p)}$ takes the form

$$f_0(t^{(p)}|t^{(1)}, \cdots, t^{(p-1)}) = \frac{C(t^{(1)}, t^{(2)}, \cdots, t^{(p)})}{\sum_z C(t^{(1)}, t^{(2)}, \cdots, t^{(p-1)}, z)}. \tag{2.8}$$

3. Testing for interaction in two 2 × 2 contingency tables

3.1. *Choice of model.* Suppose we have a 2^2 factorial experiment where the observations are "successes" or "failures." The two factors will be denoted by A and B and the factor levels by 0 and 1. The observation of the kth measurement made on the ith level of factor A and the jth level of B will be denoted by $Y_{i,j,k}$, where $i, j = 0, 1; k = 1, 2, \cdots, n_{i,j}$. Define the quantities $\theta_{i,j,k} = P\{Y_{i,j,k} = 1\}$ and $\lambda_{i,j,k} = \log \{\theta_{i,j,k}/(1 - \theta_{i,j,k})\}$. The model which we shall use for the $\lambda_{i,j,k}$ is

$$\lambda_{i,j,k} = \mu + i\alpha + j\beta + ij(\alpha\beta), \qquad i, j = 0, 1. \tag{3.1}$$

Note that this is not the usual model associated with a 2^2 experiment. If the parameters μ, α, β, and $(\alpha\beta)$ are written in terms of $\lambda_{i,j}$ (the subscript k has been dropped because $\lambda_{i,j,k}$ is constant for all k), we have

$$\begin{aligned} \mu = \lambda_{0,0}, \qquad \alpha = \lambda_{1,0} - \lambda_{0,0}, \qquad \beta = \lambda_{0,1} - \lambda_{0,0}, \\ (\alpha\beta) = (\lambda_{1,1} - \lambda_{0,1}) - (\lambda_{1,0} - \lambda_{0,0}). \end{aligned} \tag{3.2}$$

An interpretation of these parameters can be made in terms of relative risks or odds ratios. For this purpose define

$$\begin{aligned} \psi_A(j) &= \exp \{\lambda_{1,j} - \lambda_{0,j}\} = \frac{\theta_{1,j}/(1 - \theta_{1,j})}{\theta_{0,j}/(1 - \theta_{0,j})}, \qquad j = 0, 1, \\ \psi_B(i) &= \exp \{\lambda_{i,1} - \lambda_{i,0}\} = \frac{\theta_{i,1}/(1 - \theta_{i,1})}{\theta_{i,0}/(1 - \theta_{i,0})}, \qquad i = 0, 1, \end{aligned} \tag{3.3}$$

where $\theta_{i,j} = \theta_{i,j,k}$ for all k. Then we have

$$\begin{aligned} \exp \{\alpha\beta\} &= \frac{\psi_A(1)}{\psi_A(0)} = \frac{\psi_B(1)}{\psi_B(0)}, \\ \exp \{\alpha\} &= \psi_A(0), \qquad \exp \{\beta\} = \psi_B(0). \end{aligned} \tag{3.4}$$

The quantity $\exp \{\alpha\beta\}$ is simply Bartlett's definition of a second order interaction for a $2 \times 2 \times 2$ table [1]. The parameter $\psi_A(j)$ is the relative risk (odds ratio) comparing the odds ratio of success for factor A at level 1 versus 0, holding factor B at level j. An analogous interpretation holds for $\psi_B(i)$. Thus exp $\{\alpha\beta\}$ is the ratio of two relative risks (or odds ratios).

A model for $\lambda_{i,j}$ in terms of the usual main effect and interaction terms associated with a 2^2 experiment would be

$$\lambda_{i,j} = \mu + (2i - 1)(A) + (2j - 1)(B) + (2i - 1)(2j - 1)(AB), \quad i, j = 0, 1, \tag{3.5}$$

where (A) and (B) are main effect terms and (AB) is the interaction term. Clearly the relations among the parameters in the two models are

$$\begin{aligned} 2^2(A) &= -\lambda_{0,0} + \lambda_{1,0} - \lambda_{0,1} + \lambda_{1,1} = 2\alpha + (\alpha\beta), \\ 2^2(B) &= -\lambda_{0,0} - \lambda_{1,0} + \lambda_{0,1} + \lambda_{1,1} = 2\beta + (\alpha\beta), \\ 2^2(AB) &= \lambda_{0,0} - \lambda_{1,0} - \lambda_{0,1} + \lambda_{1,1} = (\alpha\beta). \end{aligned} \tag{3.6}$$

Therefore a test for H_0: $(AB) = 0$ is equivalent to H_0: $(\alpha\beta) = 0$. However tests on the main effects H_0: $(A) = 0$ or H_0: $(B) = 0$ *do not* correspond to H_0: $\alpha = 0$ or H_0: $\beta = 0$. On the other hand, tests on the null hypotheses H_0: $\alpha = 0$ and H_0: $\beta = 0$ may be regarded as tests on main effects *conditional on* $(\alpha\beta) = 0$. Hence, the parameters α and β will be referred to as *conditional main effects*. The statistical analyses are carried out by investigating the hypothesis H_0: $(\alpha\beta) = 0$. If the answer is in the affirmative, then one may carry out tests on the conditional main effects.

3.2. *Exact test for* H_0: $(\alpha\beta) = 0$. The data from the 2^2 experiment may be summarized in the two 2×2 tables depicted below. In order to avoid triple subscript notation we use the notation $s_j = \Sigma_k Y_{1,j,k}$, $r_j = \Sigma_k Y_{0,j,k}$, $t_j = r_j + s_j$, $N_j = m_j + n_j$,

(3.7)

	B_0: S	B_0: F		B_1: S	B_1: F	
A_0	r_0	$m_0 - r_0$	m_0	r_1	$m_1 - r_1$	m_1
A_1	s_0	$n_0 - s_0$	n_0	s_1	$n_1 - s_1$	n_1
	t_0	$N_0 - t_0$	N_0	t_1	$N_1 - t_1$	N_1

We shall utilize the results of Section 2 to obtain a uniformly most powerful unbiased test for H_0: $(\alpha\beta) = 0$ versus a one sided or two sided alternative. Define $\mathbf{1}_{i,j}$ as a column vector of identity elements of length $n_{i,j}$; also let $\boldsymbol{\lambda}' = (\lambda_{0,0,1}, \cdots, \lambda_{1,1,n_1})$ be the vector of the logits. The model described by equation (3.1) can then be written in matrix notation as

$$\boldsymbol{\lambda} = \begin{bmatrix} \mathbf{1}_{0,0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1}_{1,0} & \mathbf{1}_{1,0} & \mathbf{0} & \mathbf{0} \\ \mathbf{1}_{0,1} & \mathbf{0} & \mathbf{1}_{0,1} & \mathbf{0} \\ \mathbf{1}_{1,1} & \mathbf{1}_{1,1} & \mathbf{1}_{1,1} & \mathbf{1}_{1,1} \end{bmatrix} \begin{bmatrix} \mu \\ \alpha \\ \beta \\ (\alpha\beta) \end{bmatrix} \tag{3.8}$$

Let $\mathbf{Y}_{i,j}$ correspond to the $n_{i,j} \times 1$ column vector of observations $\{Y_{i,j,k}\}$ made at condition (A_i, B_j). Then the vector of sufficient statistics is

$$\text{(3.9)}\qquad \begin{bmatrix} t^{(1)} \\ t^{(2)} \\ t^{(3)} \\ t^{(4)} \end{bmatrix} = \begin{bmatrix} \mathbf{1}'_{0,0} & \mathbf{1}'_{1,0} & \mathbf{1}'_{0,1} & \mathbf{1}'_{1,1} \\ \mathbf{0} & \mathbf{1}'_{1,0} & \mathbf{0} & \mathbf{1}'_{1,1} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}'_{0,1} & \mathbf{1}'_{1,1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1}'_{1,1} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_{0,0} \\ \mathbf{Y}_{1,0} \\ \mathbf{Y}_{0,1} \\ \mathbf{Y}_{1,1} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i,j} \mathbf{1}'_{i,j}\mathbf{Y}_{i,j} \\ \sum_{j=0} \mathbf{1}'_{1,j}\mathbf{Y}_{1,j} \\ \sum_{i=0} \mathbf{1}'_{i,1}\mathbf{Y}_{i,1} \\ \mathbf{1}'_{1,1}\mathbf{Y}_{1,1} \end{bmatrix} = \begin{bmatrix} t_0 + t_1 = t \\ s_0 + s_1 = s \\ t_1 \\ s_1 \end{bmatrix}$$

Therefore to test the hypothesis H_0: $(\alpha\beta) = 0$ we require the distribution of $P\{S_1 = s_1 | T = t, S = s, T_1 = t_1\}$ which will be denoted by $f(s_1 | t, s, t_1)$. This distribution is given by equation (3.1); that is,

$$\text{(3.10)}\qquad f(s_1 | t, s, t_1) = \frac{C(t, s, t_1, s_1) \exp\{s_1(\alpha\beta)\}}{\sum_z C(t, s, t_1, z) \exp\{z(\alpha\beta)\}}.$$

The coefficient $C(t, s, t_1, s_1)$ can be found from the coefficient of $\zeta_\mu^t \zeta_\alpha^s \zeta_\beta^{t_1} \zeta_{(\alpha\beta)}^{s_1}$ in the generating function

$$\text{(3.11)}\qquad \varphi(\zeta) = (1 + \zeta_\mu)^{m_0}(1 + \zeta_\mu\zeta_\alpha)^{n_0}(1 + \zeta_\mu\zeta_\beta)^{m_1}(1 + \zeta_\mu\zeta_\alpha\zeta_\beta\zeta_{(\alpha\beta)})^{n_1}.$$

Expanding $\varphi(\zeta)$ results in

$$\text{(3.12)}\qquad \varphi(\zeta) = \sum_{i,j,k,\ell}\cdots\sum \binom{m_0}{i}\binom{n_0}{j}\binom{m_1}{k}\binom{n_1}{\ell} \zeta_\mu^{i+j+k+\ell}\zeta_\alpha^{j+\ell}\zeta_\beta^{k+\ell}\zeta_{(\alpha\beta)}^{\ell},$$

which after making the transformations

$$\text{(3.13)}\qquad t = i + j + k + \ell, \quad s = j + \ell, \quad t_1 = k + \ell, \quad s_1 = \ell,$$

results in

$$\text{(3.14)}\qquad C(t, s, t_1, s_1) = \binom{m_0}{t_0 - s_0}\binom{n_0}{s_0}\binom{m_1}{t_1 - s_1}\binom{n_1}{s_1}$$
$$= \binom{m_0}{t_0 - s + s_1}\binom{n_0}{s - s_1}\binom{n_0}{t_1 - s_1}\binom{n_1}{s_1}.$$

Note that the joint conditions $(T = t, S = s, T_1 = t_1)$ are equivalent to $T_0 = t_0$, $T_1 = t_1, S = s)$ by virtue of $T = T_1 + T_1$. Hence we shall write $f(s_1 | t, s, t_1)$ as $f(s_1 | t_0, t_1, s)$. Furthermore in what follows it will be convenient to define

$$\text{(3.15)}\qquad C(s_j, t_j) = \binom{m_j}{t_j - s_j}\binom{n_j}{s_j}, \qquad j = 0, 1.$$

Therefore

$$C(t, s, t_1, s_1) = C(s - s_1, t_0)C(s_1, t_1) \tag{3.16}$$

and we shall write (3.10) as

$$f(s_1|t_0, t_1, s) = \frac{C(s - s_1, t_0)C(s_1, t) \exp\{s_1(\alpha\beta)\}}{\sum_z C(s - z, t_0)C(z, t_1) \exp\{z(\alpha\beta)\}}. \tag{3.17}$$

The conditional distribution of S_1 when $(\alpha\beta) = 0$ is thus

$$f_0(s_1|t_0, t_1, s) = \frac{C(s - s_1, t_0)C(s_1, t_1)}{\sum_z C(s - z, t_0)C(z, t_1)}. \tag{3.18}$$

Consequently the test of significance for $H_0: (\alpha\beta) = 0$ against $H_1: (\alpha\beta) > 0$ employs the tail probability

$$P\{S_1 \geqq s_1|T_0 = t_0, T_1 = t_1, S = s\} = \sum_{w \geqq s_1} f_0(w|t_0, t_1, s). \tag{3.19}$$

The test of significance against the two sided alternative $H_1: (\alpha\beta) \neq 0$ is calculated by defining the set $W = \{w: f_0(w|t_0, t_1, s) \leqq f_0(s_1|t_0, t_1, s)\}$ and evaluating the tail probability $P = \Sigma_{w \in W} f_0(w|t_0, t_1, s)$.

3.3. *Test for main effects and decomposition of probabilities.* Analogous to the decomposition of the sums of squares associated with the general linear hypothesis for continuous type data is the decomposition of the frequency function of the observations. To see this, we note that the joint probability function is

$$\begin{aligned} &f(t, s, t_1, s_1|\mu, \alpha, \beta, (\alpha\beta)) \\ &\quad = \frac{C(t, s, t_1, s_1) \exp\{\mu t + \alpha s + \beta t_1 + (\alpha\beta)s_1\}}{\sum_i \sum_j \sum_k \sum_\ell C(i, j, k, \ell) \exp\{\mu i + \alpha j + \beta k + (\alpha\beta)\ell\}}. \end{aligned} \tag{3.20}$$

This probability function can be further decomposed into

$$\begin{aligned} &f(t, s, t_1, s_1|\mu, \alpha, \beta, (\alpha\beta)) \\ &\quad = f(s_1|t_0, t_1, s, (\alpha\beta)) f(s, t_1|t, \alpha, \beta, (\alpha\beta)) f(t|\mu, \alpha, \beta, (\alpha\beta)), \end{aligned} \tag{3.21}$$

where $f(s_1|t_0, t_1, s, (\alpha\beta))$ is given by (3.17) and

$$f(s, t_1|t, \alpha, \beta, (\alpha\beta)) = \frac{\sum_\ell C(t, s_1, t_1, \ell) \exp\{\alpha s + \beta t_1 + (\alpha\beta)\ell\}}{\sum_j \sum_k \sum_\ell C(t, j, k, \ell) \exp\{\alpha j + \beta k + (\alpha\beta)\ell\}}, \tag{3.22}$$

$$f(t|\mu, \alpha, \beta, (\alpha\beta)) = \frac{\sum_j \sum_k \sum_\ell C(t, j, k, \ell) \exp\{\mu t + \alpha j + \beta k + (\alpha\beta)\ell\}}{\sum_i \sum_j \sum_k \sum_\ell C(i, j, k, \ell) \exp\{\mu i + \alpha j + \beta k + (\alpha\beta)\ell\}}. \tag{3.23}$$

Note that $f(s, t_1 | t, \alpha, \beta, (\alpha\beta))$ is the joint distribution of (s, t_1) conditional on t and $f(t | \mu, \alpha, \beta, (\alpha\beta))$ is the marginal distribution of t. Further we can decompose

$$(3.24)\quad \begin{aligned} f(s, t_1 | t, \alpha, \beta, (\alpha\beta)) &= f(s | t_0, t_1, \alpha, (\alpha\beta)) f(t_1 | t, \alpha, \beta, (\alpha\beta)) \\ &= f(t_1 | s, t, \beta, (\alpha\beta)) f(s | t, \alpha, \beta, (\alpha\beta)), \end{aligned}$$

where

$$(3.25)\quad f(s|t_0, t_1, \alpha, (\alpha\beta)) = \frac{\sum_{s_1} C(s - s_1, t_0) C(s_1, t_1) \exp\{\alpha s + (\alpha\beta) s_1\}}{\sum_s \sum_{s_1} C(s - s_1, t_0) C(s_1, t_1) \exp\{\alpha s + (\alpha\beta) s_1\}},$$

$$(3.26)\quad f(t_1 | s, t, \beta, (\alpha\beta)) = \frac{\sum_{s_1} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\beta t_1 + (\alpha\beta) s_1\}}{\sum_{t_1} \sum_{s_1} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\beta t_1 + (\alpha\beta) s_1\}},$$

$$(3.27)\quad \begin{aligned} &f(t_1 | t, \alpha, \beta, (\alpha\beta)) \\ &= \frac{\sum_{s_1} \sum_{s} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\alpha s + \beta t_1 + (\alpha\beta) s_1\}}{\sum_{t_1} \sum_{s_1} \sum_{s} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\alpha s + \beta t_1 + (\alpha\beta) s_1\}}, \end{aligned}$$

and

$$(3.28)\quad \begin{aligned} &f(s | t, \alpha, \beta, (\alpha\beta)) \\ &= \frac{\sum_{s_1} \sum_{t_1} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\alpha s + \beta t_1 + (\alpha\beta) s_1\}}{\sum_{s} \sum_{s_1} \sum_{t_1} C(s - s_1, t - t_1) C(s_1, t_1) \exp\{\alpha s + \beta t_1 + (\alpha\beta) s_1\}}. \end{aligned}$$

If the hypothesis $H_0: (\alpha\beta) = 0$ is correct then the probability function $f(s | t_0, t_1, \alpha, (\alpha\beta) = 0) = f(s | t_0, t_1, \alpha)$ can be used to make an inference about the null hypothesis $H_0: \alpha = 0$. That is, the appropriate tail probability for the alternate hypothesis $H_1: \alpha > 0$ is

$$(3.29)\quad \sum_{z \geqq s} f(z | t_0, t_1, \alpha = 0) = \frac{\sum_{z \geqq s} \sum_{s_1} C(z - s_1, t_0) C(s_1, t_1)}{\sum_{z} \sum_{s_1} C(z - s_1, t_0) C(s_1, t_1)}.$$

The tail probability associated with the two sided alternative $H_1: \alpha \neq 0$ is

$$(3.30)\quad P = \sum_{w \in W} f(s | t_0, t_1, \alpha = 0),$$

where $W = \{w : f(w | t_0, t_1, \alpha = 0) \leqq f(s | t_0, t_1, \alpha = 0)\}$.

3.4. *Normal approximation to tests of significance.* In this section, normal approximations will be obtained for tests of significance associated with the test on the interaction and the conditional main effects. The probability of $S_j = s_j$ conditional on $t_j, j = 0, 1$, associated with a single 2×2 table is

$$(3.31)\quad p(s_j | t_j) = \frac{C(s_j, t_j) \psi_A(j)^{s_j}}{\sum_{z_j} C(z_j, t_j) \psi_A(j)^{z_j}}, \qquad j = 0, 1,$$

where $\psi_A(j)$ is the relative risk for the 2×2 table with factor B held at level j. Under the null hypothesis $H_0: \psi_A(j) = 1$, the distribution of S_j follows the hypergeometric distribution; that is,

$$p_0(s_j|t_j) = \frac{C(s_j, t_j)}{\binom{N_j}{s_j}} \tag{3.32}$$

having mean and variance

$$\begin{aligned} \mu_j &= E\{S_j|T_j = t_j\} = \frac{t_j n_j}{N_j}, \\ \sigma_j^2 &= \text{Var}\,\{S_j|T_j = t_j\} = \frac{t_j m_j n_j (N_j - t_j)}{N_j^2 (N_j - 1)}. \end{aligned} \tag{3.33}$$

Note that if $(\alpha\beta) = 0$, the distribution of S_0 conditional on $\{T_j = t_j, j = 0, 1; S = s\}$ given by (3.17) is simply

$$f(s_1|t_0, t_1, s) = \frac{p_0(s - s_1|t_0)p_0(s_1|t_1)}{\sum_z p_0(s - z|t_0)p_0(z|t_1)}. \tag{3.34}$$

For notational simplicity we shall define the random variables W_0 and W_1 by

$$P\{W_0 = s_0\} = P\{S_0 = s_0|T_0 = t_0\}, \quad P\{W_1 = s_1\} = P\{S_1 = s_1|T_1 = t_1\}.$$

When N_j is relatively large, the distribution of W_j tends to an independent normal distribution with mean u_j and variance σ_j^2. Therefore, when the normal approximation to the hypergeometric distribution holds we have

$$\begin{aligned} P\{S_1 = s_1|T_j = t_j, j = 0, 1; S_0 + S_1 = s\} &= P\{W_1 = s_1|W_0 + W_1 = s\} \\ &\cong \frac{\varphi_0(s - w_1)\varphi_1(w_1)}{\varphi(s)}, \end{aligned} \tag{3.35}$$

where $\varphi_j(x)$ is the p.d.f. of the normal distribution with parameters (μ_j, σ_j^2), and $\varphi(s)$ is the normal p.d.f. with mean $(\mu_0 + \mu_1)$ and variance $(\sigma_0^2 + \sigma_1^2)$. Substituting the appropriate p.d.f., an easy calculation shows that the approximate distribution of W_1 conditional on $W_0 + W_1 = s$ is normal with mean and variance given by

$$\begin{aligned} E\{W_1|W_0 + W_1 = s\} &= \mu_{01} = \mu_1 + \frac{\rho_{01}\sigma_1}{\sigma_0}(s - \mu_0 - \mu_1), \\ \text{Var}\,\{W_1|W_0 + W_1 = s\} &= \sigma_{01}^2 = \rho_{01}^2(\sigma_0^2 + \sigma_1^2), \end{aligned} \tag{3.36}$$

where

$$\rho_{01} = \frac{\sigma_0\sigma_1}{(\sigma_0^2 + \sigma_1^2)}. \tag{3.37}$$

Hence since the conditional distribution of W_1 is approximately $N(\mu_{01}, \sigma_{01}^2)$, a test of significance may be conducted by taking $(W_1 - \mu_{01})/\sigma_{01}$ to be a $N(0, 1)$ random variable.

If one accepts the inference that the interaction is nonexistent, then one could test the null hypotheses $H_0: \alpha = 0$ and $H_0: \beta = 0$ which refer to the conditional main effects. The appropriate conditional distributions for these hypotheses are given by (3.25) to (3.28). Under H_0, these distributions do not depend on any parameters. We shall illustrate the normal approximation for $H_0: \alpha = 0$.

Since the test of the hypothesis $H_0: \alpha = 0$ depends on the distribution of $S = S_0 + S_1$ conditional on $T_j = t_j, j = 0, 1$, we have

$$\begin{aligned} E\{S|T_0 = t_0, T_1 = t_1\} &= \mu_0 + \mu_1, \\ \text{Var}\{S|T_0 = t_0, T_1 = t_1\} &= \sigma_0^2 + \sigma_1^2. \end{aligned} \tag{3.38}$$

Consequently the large sample approximation to S, if $\alpha = 0$, is to take S to be normal with mean $\mu_0 + \mu_1$ and variance $\sigma_0^2 + \sigma_1^2$.

3.5. *Binomial approximation to nonnull distribution.* The expression for the frequency function $f(s_1|t_0, t_1, s)$, equation (3.17), associated with the test for interaction may be approximated by using the binomial approximation to the hypergeometric distribution. The hypergeometric distribution can often be approximated quite accurately by the corresponding terms of the binomial distribution which has the same mean, and as closely as possible, the same variance [5]. This approximation is

$$\frac{C(s_j, t_j)}{\binom{N_j}{t_j}} = \frac{\binom{m_j}{t_j - s_j}\binom{n_j}{s_j}}{\binom{N_j}{t_j}} \sim \binom{n_j^*}{s_j} p_j^{*s_j} q_j^{*n_j^* - s_j}, \tag{3.39}$$

where n_j^* is the nearest integer to $t_j n_j/[N_j - m_j(N_j - t_j)/(N_j - 1)]$ and $p_j^* = t_j n_j/N_j n_j$.

Using this approximation in (3.17) yields

$$f(s_1|t_0, t_1, s) \sim \frac{\binom{n_0^*}{s - s_1}\binom{n_1^*}{s_1}\{\psi^* e^{(\alpha\beta)}\}^{s_1}}{\sum_{z=z_1}^{z_2}\binom{n_0^*}{s - z}\binom{n_1^*}{z}\{\psi^* e^{(\alpha\beta)}\}^{z}}, \tag{3.40}$$

where

$$\psi^* = \frac{p_1^*/q_1^*}{p_0^*/q_0^*}, \qquad z_1 = \max(0, s - n_0^*), \qquad z_2 = \min(s, n_1^*). \tag{3.41}$$

Note that (3.40) is exactly the same expression as the frequency function associated with the nonnull conditional distribution of a single 2 × 2 contingency table. Again, employing the binomial approximation to the hypergeometric distribution; that is,

$$\frac{\binom{n_0^*}{s - s_1}\binom{n_1^*}{s_1}}{\binom{N^*}{s}} \sim \binom{\eta}{s_1}\pi^{s_1}(1 - \pi)^{\eta - s_1}, \tag{3.42}$$

where $N^* = n_0^* + n_1^*$, η = nearest integer to $sn_1^*/[N^* - n_0^*(N^* - s)/(N^* - 1)]$, and $\pi = sn_1^*/N^*\eta$, results in

$$f(s_1 | t_0, t_1, s) \sim \frac{\binom{\eta}{s_1}\{\pi\psi^* e^{(\alpha\beta)}\}^{s_1}\{1 - \pi\}^{\eta - s_1}}{[\pi\psi^* e^{(\alpha\beta)} + 1 - \pi]^\eta} = \binom{\eta}{s_1} P^{s_1} Q^{\eta - s_1}, \tag{3.43}$$

where

$$P = \frac{\pi\psi^* e^{(\alpha\beta)}}{[\pi\psi^* e^{(\alpha\beta)} + 1 - \pi]}, \qquad Q = 1 - P. \tag{3.44}$$

Consequently, an approximation to $f(s_1 | t_0, t_1, s)$ is to take the random variable S_1 to be approximately distributed as a binomial random variable with sample size η and success probability P. Hence a confidence interval on P can, by suitable transformation, be made into a confidence interval on $e^{(\alpha\beta)}$. That is, if (P_1, P_2) are $100(1 - 2\alpha)$ per cent confidence limits on P, then

$$\frac{(1 - \pi)/(1 - P_i)}{\psi^*(\pi/P_i)}, \qquad i = 1, 2, \tag{3.45}$$

are approximate confidence limits for $e^{(\alpha\beta)}$. Furthermore, another approximation of the significance test for the null hypothesis H_0: $(\alpha\beta) = 0$ versus H_1: $(\alpha\beta) > 0$ is to compute the tail area probabilities

$$P\{S_1 \geqq s_1 | (\alpha\beta) = 0\} = \sum_{k=s_1}^{\eta} \binom{\eta}{k} P_0^k Q_0^{\eta - k}, \tag{3.46}$$

where $P_0 = \pi\psi^*/[\pi\psi^* + 1 - \pi]$.

4. The general case of n factors

4.1. *Preliminaries and notation.* In this section we extend the treatment of the analysis of 2×2 contingency tables for the situation where the number of tables is a power of two. It is convenient to use the notation associated with factorial experiments. Let $A_1, A_2, \cdots, A_n$ represent n factors each at two levels. There will then be 2^n factorial combinations. If we fix on one factor, say A_1, this situation may be regarded as having 2^{n-1} 2×2 contingency tables involving the two levels of A_1 where each contingency table represents a fixed combination of the remaining $(n - 1)$ factors.

The general development is eased if one adopts an operational calculus suited to factorial experiments (see Kurkjian and Zelen [8]). Let a factorial combination be denoted by the n tuple binary number $x = (x_1, x_2, \cdots, x_n)$ where $x_i = 0$ or 1, $i = 1, 2, \cdots, n$, depending on whether the ith factor is at the "low" level or "high" level, respectively. Throughout this section, we will have need for ordering the 2^n n digit binary numbers in a standard order. For this purpose we use the operation of the symbolic direct product which is designated by $\otimes$ (see [8]). Define the vector δ by $\delta' = (0, 1)$. Then the standard order for $n = 2$ (four treatment combinations) is given by the rows of $\delta \otimes \delta$ which is

$$\delta \otimes \delta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{4.1}$$

The standard order for $n = 3$ is given by the rows of

$$\delta \otimes \delta \otimes \delta = \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}. \tag{4.2}$$

The generalization to arbitrary n is clear.

Let $Y(x)$ denote a binary random variable representing the outcome of the xth treatment combination. Also define

$$\begin{aligned} &\theta(x) = P\{Y(x) = 1\}, \qquad 1 - \theta(x) = P\{Y(x) = 0\}, \\ &\lambda(x) = \log \{\theta(x)/(1 - \theta(x)\}. \end{aligned} \tag{4.3}$$

The generalized interaction among the factors $A_{i_1}, A_{i_2}, \cdots, A_{i_p}$ will be denoted by $a_{i_1 i_2 \cdots i_p}$. Another way of designating this generalized interaction is to define

$$\begin{aligned} &z_i = \begin{cases} 1 & \text{if factor } A_i \text{ is included in the generalized interaction,} \\ 0 & \text{otherwise,} \end{cases} \\ &z = (z_1, z_2, \cdots, z_n), \end{aligned} \tag{4.4}$$

and let $a(z) = a_{i_1 i_2 \cdots i_p}$. That is, the generalized interaction can also be designated by an n digit binary number. The binary number $(0, 0, \cdots, 0)$ will refer to a constant term; that is, $a(0, 0, \cdots, 0) = \mu$.

For the purpose of writing $\lambda(x)$ as a function of the generalized interactions, define

$$\delta^{z_i} = \begin{cases} (0, 1)' & \text{if } z_i = 1, \\ (1, 1)' & \text{if } z_i = 0, \end{cases} \tag{4.5}$$

and

$$\delta^z = (\delta_1^{z_1} \times \delta_2^{z_2} \times \cdots \times \delta_n^{z_n}), \tag{4.6}$$

where $\times$ denotes the Kronecker product. Then if λ is the vector of $\{\lambda(x)\}$ arranged in the standard order we have

$$\lambda = \sum_z \delta^z a(z), \tag{4.7}$$

where the summation is over all n digit binary numbers. The $\lambda(x)$ corresponding to a particular treatment combination can then be written as

$$\begin{aligned} \lambda(x) &= \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix}\right]' \lambda \\ &= \sum_z \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix}' \delta^{z_1} \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix}' \delta^{z_2} \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix}' \delta^{z_n}\right] a(z), \end{aligned} \tag{4.8}$$

where $\times$ denotes Kronecker product multiplication. It is easy to verify that

$$\begin{pmatrix} 1 - x_i \\ x_i \end{pmatrix}' \delta^{z_i} = 1 - z_i(1 - x_i), \qquad i = 1, 2, \cdots, n. \tag{4.9}$$

Hence we have

$$\lambda(x) = \sum_z \prod_{i=1}^n [1 - z_i(1 - x_i)]a(z) = \sum_z \varphi(x, z)a(z), \tag{4.10}$$

where

$$\varphi(x, z) = \prod_{i=1}^n [1 - z_i(1 - x_i)].$$

4.2 *The logistic model.* At condition x let

$$\theta(x) = P\{Y(x) = 1\} = \frac{\exp\{\lambda(x)\}}{1 + \exp\{\lambda(x)\}}. \tag{4.11}$$

Then if $Y_1(x), Y_2(x), \cdots, Y_{m(x)}(x)$ denote $m(x)$ independent binary random variables made at treatment x, we have

(4.12)

$$\begin{aligned} P\{\mathbf{Y}(x) = \mathbf{y}(x)\} &= P\{Y_1(x) = y_1(x), Y_2(x) = y_2(x), \cdots, Y_{m(x)}(x) = y_{m(x)}(x)\} \\ &= \frac{\exp\{s(x)\lambda(x)\}}{[1 + \exp\{\lambda(x)\}]^{m(x)}}, \end{aligned}$$

where $s(x) = \Sigma_{j=1}^{m(x)} y_j(x)$ is the total number of one's at condition x. Hence if $\mathbf{Y}$ denotes the vector of $\{\mathbf{Y}(x)\}$ over all treatment combinations and $\mathbf{y}$ is the vector of outcomes we have

$$P\{\mathbf{Y} = \mathbf{y}\} = \prod_x \frac{\exp\{s(x)\lambda(x)\}}{[1 + \exp\{s(x)\lambda(x)\}]^{m(x)}} = \frac{\exp \sum_x s(x)\lambda(x)}{\prod_x [1 + \exp\{\lambda(x)\}]^{m(x)}}, \tag{4.13}$$

where all the products and sums are taken over the n digit binary numbers x.

Since $\sum_x s(x)\lambda(x) = \sum_x s(x) \sum_z \varphi(x, z)a(z) = \sum_z t(z)a(z)$ where

$$t(z) = \sum_x \varphi(x, z)s(x), \tag{4.14}$$

we have

$$P\{\mathbf{Y} = \mathbf{y}\} = \frac{\exp \sum_z t(z)a(z)}{\prod_x [1 + \exp \sum_z \varphi(x, z)a(x)]^{m(x)}}. \tag{4.15}$$

Thus we have shown that $\{t(z)\}$ is jointly sufficient for the 2^n parameters $\{a(z)\}$.

Since a set of sufficient statistics exists which is actually a minimal set, we can reduce the distribution of $\mathbf{Y}$ given by (4.13) to the joint distribution of the sufficient statistics $\{t(x)\}$. Consequently, if $\mathbf{t}$ is the vector of sufficient statistics, we have

$$p(\mathbf{t}) = P\{\mathbf{T} = \mathbf{t}\} = \frac{C(\mathbf{t}) \exp \sum_z t(z)a(z)}{\prod_x [1 + \exp \sum_z \varphi(x, z)a(z)]^{m(x)}}. \tag{4.16}$$

where $C(\mathbf{t})$ is a quantity dependent on $\mathbf{t}$ and *not* on $\{a(z)\}$ which makes $\Sigma_t\, p(\mathbf{t}) = 1$.

In order to find $C(\mathbf{t})$, note that

$$\sum_{\mathbf{t}} C(\mathbf{t}) \exp \sum_z t(z)a(z) = \prod_x [1 + \exp \sum_z \varphi(x, z)a(z)]^{m(x)}. \tag{4.17}$$

Letting $\xi(z) = \exp\{a(z)\}$ results in (4.17) being written as

$$\sum_{\mathbf{t}} C(\mathbf{t}) \prod_z \xi(z)^{t(z)} = \prod_x [1 + \prod_z \xi(z)^{\phi(x,z)}]^{m(x)}. \tag{4.18}$$

Thus

$$\Phi(\boldsymbol{\xi}) = \prod_x [1 + \prod_z \xi(z)^{\phi(x,z)}]^{m(x)} \tag{4.19}$$

is the generating function which enables the coefficients $C(\mathbf{t})$ to be found.

Expanding $\Phi(\xi)$ gives

$$\Phi(\boldsymbol{\xi}) = \prod_x \sum_{r(x)=0}^{m(x)} \binom{m(x)}{r(x)} \prod_z \xi(z)^{\Sigma_x \phi(x,z)r(x)} \tag{4.20}$$

and therefore setting

$$t(z) = \sum_x \varphi(x, z)r(x), \tag{4.21}$$

we have

$$C(\mathbf{t}) = \prod_x \binom{m(x)}{r(x)}, \tag{4.22}$$

where the $\{r(x)\}$ in (4.22) are replaced by solving the 2^n simultaneous equations (4.21) for $\{r(x)\}$.

We shall now solve the system of linear equations (4.21) for $\{r(x)\}$. Writing $t(z)$ as

$$(4.23)\qquad t(z) = \sum_x \varphi(x, z) r(x) = \sum_x \prod_z \begin{pmatrix} 1 - x_i \\ x_i \end{pmatrix}' \delta^z r(x)$$

$$= [\delta^z]' \sum_x \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix} \right] r(x)$$

and substituting

$$(4.24)\qquad r(x) = \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix} \right]' \mathbf{r}$$

where $\mathbf{r}$ is the column vector of $\{r(x)\}$ arranged in standard order, results in

$$(4.25)\qquad t(z) = [\delta^z]' \sum_x \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} (1 - x_1, x_1) \right.$$

$$\left. \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} (1 - x_2, x_2) \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix} (1 - x_n, x_n) \right] \mathbf{r}.$$

Since

$$(4.26)\qquad \begin{pmatrix} 1 - x_i \\ x_i \end{pmatrix} (1 - x_i, x_i) = \begin{cases} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} & \text{if } x_i = 0, \\ \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} & \text{if } x_i = 1, \end{cases}$$

the matrix of order 2^n

$$(4.27)\qquad \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} (1 - x_1, x_1) \right.$$

$$\left. \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} (1 - x_2, x_2) \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix} (1 - x_n, x_n) \right]$$

consists of all zeros except for a single entry of unity on the main diagonal which is in the same position as $x = (x_1, x_2, \cdots, x_n)$ is in the standard order. Therefore

$$(4.28)\qquad \sum_x \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} (1 - x_1, x_1) \right.$$

$$\left. \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} (1 - x_2, x_2) \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix} \right] = I$$

and thus

$$(4.29)\qquad t(z) = [\delta^z]' \mathbf{r}.$$

Now if the matrix M is defined by

$$M = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \tag{4.30}$$

the solution of (4.29) is

$$\mathbf{r} = [M \times M \times \cdots \times M]\mathbf{t}. \tag{4.31}$$

The proof is immediate by substituting (4.31) in (4.29); that is,

$$\begin{aligned} t(z) &= [\delta^z]'[M \times M \times \cdots \times M]\mathbf{t} \\ &= \left[\begin{pmatrix} 1 - z_1 \\ z_1 \end{pmatrix} \times \begin{pmatrix} 1 - z_2 \\ z_2 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - z_n \\ z_n \end{pmatrix}\right]' \mathbf{t} = t(z) \end{aligned} \tag{4.32}$$

as $(\delta^{z_i})'M = (1 - z_i, z_i)$. Furthermore, since

$$r(x) = \left[\begin{pmatrix} 1 - x_1 \\ x_1 \end{pmatrix} \times \begin{pmatrix} 1 - x_2 \\ x_2 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - x_n \\ x_n \end{pmatrix}\right]' \mathbf{r}, \tag{4.33}$$

we have

$$r(x) = \left[\begin{pmatrix} 1 - x_1 \\ 2x_1 - 1 \end{pmatrix} \times \begin{pmatrix} 1 - x_2 \\ 2x_2 - 1 \end{pmatrix} \times \cdots \times \begin{pmatrix} 1 - x_n \\ 2x_n - 1 \end{pmatrix}\right]' \mathbf{t} \tag{4.34}$$

by virtue of

$$(1 - x_i, x_i)M = (1 - x_i, 2x_i - 1). \tag{4.35}$$

Thus, we have shown that the coefficient $C(\mathbf{t})$ in (4.16) is given explicitly by $C(\mathbf{t}) = \Pi_x\binom{m(x)}{r(x)}$ where $r(x)$ is found from (4.34).

Another way of viewing this analysis problem is to consider the data arranged in 2^{n-1} 2×2 contingency tables where each table records the number of successes and failures for factor A_1 at its two levels. The 2^{n-1} different tables correspond to all possible combinations of the remaining factors $A_2, A_3, \cdots, A_n$. An $(n - 1)$ digit binary number, say $y = (y_2, y_3, \cdots, y_n)$, can be used to denote a particular combination of the $(n - 1)$ factors. It may also be convenient to number the 2^{n-1} contingency tables in the base ten number system. For this purpose let the ith table, corresponding to the combination $y = (y_2, y_3, \cdots, y_n)$, be

$$i = y_2 2^{n-2} + y_3 2^{n-3} + \cdots + y_n. \tag{4.36}$$

where $i = 0, 1, \cdots, N$ $(N = 2^{n-1} - 1)$. Then from (4.14) we have

$$\begin{aligned} t(0, y) &= \sum_x \varphi(x, 0, y)s(x) \\ &= \sum_x \varphi(x, 0, y)[s(0, x_2, \cdots, x_n) + s(1, x_2, \cdots, x_n)], \\ t(1, y) &= \sum_x \varphi(x, 1, y)s(x) = \sum_x \varphi(x, 1, y)[s(1, x_2, \cdots, x_n)]. \end{aligned} \tag{4.37}$$

Thus we have that an equivalent set of sufficient statistics are $\{s(0, y) + s(1, y)\}$ and $\{s(1, y)\}$ where y ranges over all $(n - 1)$ digit binary numbers. It will be convenient to let

(4.38)
$$\begin{aligned} t_i &= s(0, y) + s(1, y) = \text{total number of successes in the } i\text{th contingency table},\\ s_i &= s(1, y) = \text{total successes for level one of factor } A_1 \text{ in the } i\text{th contingency table},\\ m_i &= m(0, y) = \text{total number of trials in the } i\text{th contingency table for level zero of factor } A_1,\\ n_i &= m(1, y) = \text{total number of trials in the } i\text{th contingency table for level one of factor } A_1. \end{aligned}$$

Using the above change of notation, $C(\mathbf{t})$ may be written as

(4.39)
$$C(\mathbf{t}) = \prod_{i=0}^{N} C(s_i, t_i), \qquad N = 2^{n-1} - 1,$$

where

(4.40)
$$C(s_i, t_i) = \binom{m_i}{t_i - s_i}\binom{n_i}{s_i}.$$

Note that $t(1, y)$ (4.37) can be written as

(4.41)
$$t(1, y) = [\delta_2^{y_2} \times \cdots \times \delta_n^{y_n}]' \mathbf{S}_1,$$

where $\mathbf{S}_1' = (s_0, s_1, \cdots, s_N)$. However (4.41) is the same form as (4.34). Consequently, the solution of the $\mathbf{S}_1$ vector is

(4.42)
$$\mathbf{S}_1 = [M \times \cdots \times M]\mathbf{t}_1,$$

where $\mathbf{t}_1$ is the vector of $\{t(1, y)\}$. Also the analogue of (4.34) is

(4.43)
$$s_i = s(1, y) = \left[\binom{1 - y_2}{2y_2 - 1} \times \binom{1 - y_3}{2y_3 - 1} \times \cdots \times \binom{1 - y_n}{2y_n - 1}\right]' \mathbf{t}_1.$$

Similarly $t(0, y)$ (4.37) may be written as

(4.44)
$$\mathbf{t}_0 = [\delta_2^{y_2} \times \delta_2^{y_3} \times \cdots \times \delta_n^{y_n}]' \mathbf{t},$$

where $\mathbf{t}_0$ is the vector of $\{t(0, y)\}$ and $\mathbf{t}' = (t_0, t_1, \cdots, t_N)$. Hence $\mathbf{t} = [M \times M \times \cdots \times M]\mathbf{t}_0$.

4.3. *Conditional test of significance for interactions.* The analysis of the set of the 2^{n-1} contingency tables proceeds by testing the highest order interaction. If the inference is made that this interaction exists, the 2^{n-1} contingency tables are partitioned into two sets, each of 2^{n-2} tables each. An independent analysis is done on each set, first testing for the highest order interaction. On the other hand, if the inference is made that the n factor interaction is zero, the next step in the analysis is to make inferences on the n interactions involving $(n - 1)$

factors, assuming the n factor interaction is zero. The analysis proceeds in this way, partitioning the set of tables whenever the highest order interaction is real and testing the next lowest order interactions if the highest order interaction is negligible. In this section we exhibit the appropriate tests of significance for carrying out the necessary significance tests. The significance tests are all based on conditional reference sets and are parameter free in the same sense as the Fisher-Irwin analysis of the 2×2 contingency table.

Let x be a given n digit binary number and let $\bar{\mathbf{t}}(x)$ denote the $2^n - 1$ vector of the $\{t(z)\}$ excluding $t(x)$. Then we can write $\mathbf{t}$ as $\mathbf{t} = (\bar{\mathbf{t}}(x), t(x))$. Using (4.16), we have

$$(4.45) \qquad p(t(x)|\bar{\mathbf{t}}(x)) = P\{T(x) = t(x)|\,\bar{\mathbf{T}}(x) = \bar{\mathbf{t}}(x)\}$$
$$= \frac{C(\bar{\mathbf{t}}(x), t(x)) \exp \{t(x)a(x)\}}{\sum_{t(x)} C(\bar{\mathbf{t}}(x), t(x)) \exp \{t(x)a(x)\}},$$

where the summation in the denominator is over the range of $t(x)$. When the hypothesis $H_0: a(x) = 0$ is true, then (4.45) becomes

$$(4.46) \qquad p_0(t(x)|\bar{\mathbf{t}}(x)) = \frac{C(\bar{\mathbf{t}}(x), t(x))}{\sum_{t(x)} C(\bar{\mathbf{t}}(x), t(x))}.$$

Thus $p_0(t(x)|\bar{\mathbf{t}}(x))$ can be used to carry out a test of significance for the null hypothesis $H_0: a(x) = 0$. The test may be one sided or two sided depending on the alternative hypothesis.

The test for the highest order interaction corresponds to taking $x = (1, 1, \cdots, 1)$. Using this value in (4.14) we find that

$$(4.47) \qquad t(1, 1, \cdots, 1) = s(1, 1, \cdots, 1) = \text{total number of ones (or successes) at factorial combination where all factors are at the upper level.}$$

Hence, the appropriate distributions for carrying out inferences on $a(1, 1, \cdots, 1)$ are (4.45) and (4.46).

If one concludes that the n factor interaction is zero, then the next step in the analysis is to make an inference on the interactions involving $(n - 1)$ factors conditional on the n factor interaction being zero. Let $\mathbf{1}$ be a row vector having n elements. Then the marginal distribution of $\bar{\mathbf{T}}(\mathbf{1})$ may be written

$$(4.48) \qquad P\{\bar{\mathbf{T}}(\mathbf{1}) = \bar{\mathbf{t}}(\mathbf{1})\} = \frac{\sum_{t(\mathbf{1})} C(\bar{\mathbf{t}}(\mathbf{1}), t(\mathbf{1})) \exp \sum_z t(z)a(z)}{\prod_x [1 + \exp \sum_z \varphi(x, z)a(z)]^{m(x)}},$$

and the conditional distribution of $T(x)$ (corresponding to $a(x)$) is

$$(4.49) \qquad P\{T(x) = t(x)|\,\bar{\mathbf{T}}(x) = \bar{\mathbf{t}}(x) \text{ excluding } T(\mathbf{1}) = t(\mathbf{1})\}$$
$$= \frac{\{\sum_{t(\mathbf{1})} C(\bar{\mathbf{t}}(\mathbf{1}), t(\mathbf{1}))\} \exp \{t(x)a(x)\}}{\sum_{t(x)} \{\sum_{t(\mathbf{1})} C(\bar{\mathbf{t}}(\mathbf{1}), t(\mathbf{1}))\} \exp \{t(x)a(x)\}}.$$

The distribution under the null hypothesis $H_0: a(x) = 0$ is obtained from (4.49) by setting $a(x) = 0$.

The general procedure for making inferences on lower order interactions is clear. A test on an interaction among p factors is only carried out if all higher order interactions involving the p factors are assumed to be zero. That is, define:

(4.50) $\mathscr{A}_p = \left\{x : a(x) = 0, \sum_{i=1}^{n} x_i > p\right\}$ = subset of all n digit binary numbers associated with interactions involving at least $(p + 1)$ factors;

(4.51) $\bar{\mathscr{A}}_p$ = complement of $\mathscr{A}_p$;

(4.52) $\mathscr{B}(a(x)) = \{x_i : x_i = 1, a(x)\}$ = set of x_i elements for which $x_i = 1$ in $a(x)$;

(4.53) $\mathscr{C}_p = \bigcap_{x \in \mathscr{A}} \mathscr{B}(a(x))$ = subset of x_i elements which are equal to unity in $\mathscr{A}_p$.

Then if $a(x)$ corresponds to an interaction involving p factors such that all $x_i = 1$ in $a(x)$ belong to $\mathscr{C}_p$ we have

$$(4.54) \quad P\{T(x) = t(x) \mid T(y) = t(y), y \neq x \text{ and all } y \in \bar{\mathscr{A}}_p\} = \frac{\{\sum \cdots \sum_{t(z)} C(\mathbf{t})\} \exp t(x)a(x)}{\sum_{t(x)} \{\sum \cdots \sum_{t(z)} C(\mathbf{t})\} \exp t(x)a(x)},$$

where the summation in brackets in both numerator and denominator is over the range of $t(z)$ for *all* $z \in \mathscr{A}_p$.

4.4 *The analysis for four* 2×2 *tables* $(n = 3)$. In this section we illustrate how special cases can easily be obtained from the general results of the preceeding sections. We shall take the case $n = 3$ corresponding to four 2×2 tables. The first step in the analysis is to obtain the $C(\mathbf{t})$ coefficients using (4.40) and (4.41). The identification between the binary and base ten notation for two digit numbers is given in Table I.

TABLE I

IDENTIFICATION BETWEEN BINARY AND BASE TEN NOTATION

Binary (y)	Base Ten (i)
(0, 0)	0
(0, 1)	1
(1, 0)	2
(1, 1)	3

Let the four tables be identified with the above indices. Then if s_i is the total success for level one of A_1 and t_i is the total number of successes for table i, $i = 0, 1, 2, 3$. We easily calculate from (4.41) and (4.43),

$$(4.55)\qquad t_1(0) = t(1,0,0) = \sum_{i=0}^{3} s_i,\ t_1(2) = t(1,1,0) = s_2 + s_3,$$

$$t_1(1) = t(1,0,1) = s_1 + s_3,\ t_1(3) = t(1,1,1) = s_3,$$

and

$$(4.56)\qquad \begin{aligned} s_0 &= t_0(1) - t_1(1) - t_1(2) + t_1(3), \\ s_1 &= t_1(1) - t_1(3), \\ s_2 &= t_2(1) - t_1(3), \\ s_3 &= t_1(3). \end{aligned}$$

Therefore, from (4.39) and (4.40) we have

$$(4.57)\qquad C(\mathbf{t}) = \prod_{i=0}^{3} C(s_i, t_i),$$

where (4.56) gives the values of s_i in terms of $t_1(i)$. Thus, the conditional probability distribution associated with the test of the highest order interaction is

$$(4.58)\qquad P\left\{T(1,1,1) = s_3 \,|\, T_i = t_i (i = 0,1,2,3), S_2 + S_3 = t_1(2),\right.$$
$$\left. S_1 + S_3 = t_1(1), \sum_{i=0}^{3} S_i = t_1(0)\right\}$$
$$= \frac{\prod_{i=0}^{3} C(s_i, t_i) \exp\{s_3 a(1,1,1)\}}{\sum_{s_3} \prod_{i=0}^{3} C(s_i, t_i) \exp\{s_3 a(1,1,1)\}}.$$

Tail area probabilities for H_0: $a(1,1,1) = 0$ may be calculated from (4.58) by setting $a(1,1,1) = 0$.

If one concludes that the $a(1,1,1)$ interaction is zero, the next step in the analysis is to make an inference on the conditional interactions involving two factors; that is, $a(1,1,0)$, $a(1,0,1)$, and $a(0,1,1)$. We shall illustrate the appropriate conditional test for the $a(1,1,0)$ and $a(0,1,1)$ interactions. Using (4.49) with $x = (1,1,0)$, we have

$$(4.59)\qquad P\left\{T_1(2) = t_1(2) \,|\, T_i = t_i (i = 0,1,2,3), S_1 + S_3 = t_1(1), \sum_{i=0}^{3} S_i = t_1(0)\right\}$$
$$= \frac{\left\{\sum_{s_3} \prod_{i=0}^{3} C(s_i, t_i)\right\} \exp\{t_1(2) a(1,1,0)\}}{\sum_{t_1(2)} \left\{\sum_{s_3} \prod_{i=0}^{3} C(s_i, t_i)\right\} \exp\{t_1(2) a(1,1,0)\}}.$$

Note that both the inference on $a(1, 1, 1)$ and $a(1, 1, 0)$ were made conditional on $T_i = t_i$, $i = 0, 1, 2, 3$. On the other hand, the inference for $a(0, 1, 1)$ is conditional on the $S_i = s_i$, $i = 0, 1, 2, 3$. The appropriate conditional test for $H_0: a(0, 1, 1) = 0$ is obtained by finding the probability of $T(0, 1, 1) = t(0, 1, 1)$ conditional on $S_i = s_i$, $i = 0, 1, 2, 3$, $T(0, y) = t(0, y)$ for $y = (0, 0), (0, 1), (1, 0)$.

From (4.37) we have

$$\begin{aligned} t_0(0) &= t(0, 0, 0) = \sum_{i=0}^{3} t_i, & t_0(2) &= t(0, 1, 0) = t_2 + t_3 \\ t_0(1) &= t(0, 0, 1) = t_1 + t_3, & t_0(3) &= t(0, 1, 1) = t_3. \end{aligned} \tag{4.60}$$

Also, solving for the $\{t_i\}$ terms of $\{t_0(i)\}$ results in

$$\begin{aligned} t_0 &= t_0(0) - t_0(1) - t_0(2) + t_0(3), \\ t_1 &= t_0(1) - t_0(3), \\ t_2 &= t_0(2) - t_0(3), \\ t_3 &= t_0(3). \end{aligned} \tag{4.61}$$

The conditional distribution of T_3 is

$$\begin{aligned} P\Big\{T_3 = t_3 \,\big|\, S_i = s_i (i = 0, 1, 2), \sum_{i=0}^{3} T_i = t_0(0), \\ T_1 + T_3 = t_0(1), T_2 + T_3 = t_0(2)\Big\} \\ = \frac{\sum_{s_3} \prod_{i=0}^{3} C(s_i, t_i) \exp\{t_3 a(0, 1, 1)\}}{\sum_{t_3}\sum_{s_3} \prod_{i=0}^{3} C(s_i, t_i) \exp\{t_3 a(0, 1, 1)\}}, \end{aligned} \tag{4.62}$$

where the t_i is replaced in the $\{C(s_i, t_i)\}$ by (4.61). The test of significance is obtained by setting $a(0, 1, 1) = 0$ in (4.62) and calculating the appropriate tail probability.

◇ ◇ ◇ ◇ ◇

Added in proof. The recently published book by D. R. Cox, *Analysis of Binary Data*, London, Methuen, 1970, summarizes much of the results of [2] and [3] as well as several generalizations.

REFERENCES

[1] M. S. BARTLETT, "Contingency table interactions," *J. Roy. Statist. Soc. Supplemnt*, Vol. 2 (1935), pp. 248–252.

[2] D. R. COX, "The regression analysis of binary sequences," *J. Roy. Statist. Soc. Ser. B*, Vol. 20 (1958), pp. 215–242.

[3] ———. "Some procedures connected with the logistic qualitative response curve." *Research Papers in Statistics* (edited by F. N. David). New York, Wiley, 1966.

[4] L. A. GOODMAN, "The multivariate analysis of qualitative data: interactions among multiple classifications," *J. Amer. Statist. Assoc.*, Vol. 65 (1970), pp. 226–256.

[5] HARVARD UNIVERSITY, *Tables of the Cumulative Binomial Probability Distribution*, Cambridge, Harvard University Press, 1955.

[6] H. H. KU and S. KULLBACK, "Interaction in multidimensional contingency tables: an information theoretic approach," *J. Res. Nat. Bur. Standards Section B*, Vol. 72 (1968), pp. 159–199.

[7] S. KULLBACK, M. KUPPERMAN, and H. H. KU, "An application of information theory to the analysis of contingency tables, with a table of $2n \log n$, $n = 1(1)10{,}000$," *J. Res. Nat. Bur. Standards Section B*, Vol. 66 (1962), pp. 217–243.

[8] B. KURKJIAN and M. ZELEN, "A calculus for factorial arrangements," *Ann. Math. Statist.*, Vol. 33 (1962), pp. 600–619.

[9] E. L. LEHMANN, *Testing Statistical Hypotheses*, New York, Wiley, 1959. Chapter 4.

[10] B. N. LEWIS, "On the analysis of interaction in multidimensional contingency tables," *J. Roy. Statist. Soc. Ser. A*, Vol. 125 (1962), pp. 88–117.

[11] R. L. PLACKETT, "Multidimensional contingency tables: a survey of models and methods," *Proceedings of the 37th Session, Bulletin of the I.S.I.*, Vol. 43, Book 1 (1969), pp. 133–142.