The Annals of Statistics
1983, Vol. 11, No. 3, 724–738

ADDITIVE AND MULTIPLICATIVE MODELS AND INTERACTIONS

BY J. N. DARROCH AND T. P. SPEED

The Flinders University of South Australia and C.S.I.R.O.

A unified treatment is given of the classical additive models for complete factorial experiments and of multiplicative models and Lancaster-additive models for multi-dimensional contingency tables. The models are characterised by properties of being simplest subject to having a prescribed set of marginals. It is shown that, by using averaging operators and the notion of a generalised interaction, the interaction properties of these models can be derived very simply.

1. Introduction. Interaction models provide simplified structures for the arrays of unknown parameters which arise in *factorial experiments* and in multidimensional *contingency tables*. These two fields of application will be considered side by side, rather more attention being given to contingency tables.

In a factorial experiment there are, say, s factors $A_1, \cdots, A_s$ and a single response y. If the factors have $r_1, \cdots, r_s$ levels there are $r_1 \times \cdots \times r_s$ different combinations of levels called *cells*. The expected value $\mathbb{E}y = \eta$ of the response varies from cell to cell and inferential attention is focused on the array η of the $r_1 \times \cdots \times r_s$ values of η.

In a *pure response* s-dimensional contingency table there are s categorical variables $X_1, \cdots, X_s$ taking $r_1, \cdots, r_s$ values. This time the unknown parameter at each cell is the probability p of that particular combination of response values. The following discussion also applies to s-dimensional contingency tables in which some of the dimensions correspond to factors and the remainder to responses. The probability p is then the probability of the response values given the factor levels. There remains one further model for contingency tables. In it the $r_1 \times \cdots \times r_s$ frequencies are independent Poisson variables and the theory of this paper is applied to the array of their mean values μ.

The standard models for η, p, μ or some function of them are defined by *linear subspaces* of $\mathbb{R}^{\mathscr{I}}$, where

$$\mathscr{I} = \{i = (i_1, \cdots, i_s) : 1 \le i_\sigma \le r_\sigma, \sigma = 1, \cdots, s\}.$$

They are usually obtained by introducing a system of *interactions* and then requiring that a subset of these interactions vanish. This may be quite appropriate with additive models for factorial experiments, where the individual interactions can have a practical interpretation, but it is not necessarily so with multiplicative models for contingency tables. One of the aims of this paper is to give a simple account of an alternative approach in which we define models first (Section 2) and interactions later (Section 4). In doing so we take the opportunity to compare and contrast additive and multiplicative models, and to note the similarities and differences between two widely used parametrizations.

There is a certain amount of overlap in subject matter between this paper and the work of Haberman (1974, 1975) but the mathematical treatments of the common material are substantially different. Andersen (1975) gives a very clear summary of the general properties of interaction subspaces, applicable either to additive or multiplicative models, whilst other general treatments are by Mann (1949), Good (1958, 1963), Kurkjian and Zelen (1962), Grizzle, Starmer and Koch (1969), Goodman (1970) and Davidson (1973). Writings which concentrate upon multiplicative models for probabilities include several books: Haberman (1974, 1978, 1979), Bishop, Fienberg and Holland (1975), Fienberg (1977),

Received February 1980; revised December 1982.
AMS 1980 *subject classifications*. 62E10, 62E30.
Key words and phrases. Generalised interaction, Lancaster model, linear model log-linear model.

Gokhale and Kullback (1978), and Plackett (1981). Lancaster's theory of interaction and generalised correlation can be found in his book (1969), although the formulation given here (for finitely-valued random variables) is slightly different from his, being chosen to facilitate comparisons with other models. Further literature references are given in the body of the paper.

Inference matters are not discussed apart from a few comments on least-squares, sufficient reductions and maximum likelihood estimation. There is also no discussion of experimental design questions. A number of the results in this paper are new but in general the emphasis is on unifying existing results and on proving them by elementary methods.

2. Models and marginals. In this section we introduce the models which will be the main topic of the paper. The s factors or responses will be labeled by elements of $S = \{1, 2, \cdots, s\}$, subsets of which will be denoted by a, b, c, d. As in the introduction $\sigma \in S$ is supposed to have r_σ values (levels or response categories), and we write $\mathscr{I}$ for the set of cells i; precisely $\mathscr{I} = \{i = (i_\sigma) : i \le i_\sigma \le r_\sigma, \sigma \in S\}$. More generally we write i_a for the sub-tuple $i_a = (i_\sigma : \sigma \in a)$, $a \subseteq S$.

2.1 *The models.* Let $\mathscr{A}$ be a collection of subsets of S. The linear subspace $\Omega_{\mathscr{A}}$ of $\Omega = \mathbb{R}^{\mathscr{I}}$ is defined by the property that the function $f = (f(i) : i \in \mathscr{I})$ belongs to $\Omega_{\mathscr{A}}$ if and only if

$$f(i) = \textstyle\sum_{a \in \mathscr{A}} \lambda_a(i_a) \tag{2.1}$$

for some functions $\{\lambda_a : a \in \mathscr{A}\}$. Having defined $\Omega_{\mathscr{A}}$, the *model* $M_{\mathscr{A}}$ for f is simply the property that f belongs to $\Omega_{\mathscr{A}}$. The collection $\mathscr{A}$ is called the *generating class* of the model. Given $\mathscr{A}$ let $\mathscr{A}^*$ denote the sub-collection of elements of $\mathscr{A}$ which are maximal with respect to inclusion. It is clear that $M_{\mathscr{A}^*}$ is the same model as $M_{\mathscr{A}}$ because if $b \subseteq a$, then $\lambda_a(i_a) + \lambda_b(i_b) = \mu_a(i_a)$. Whilst it is economical in practice to work with $\mathscr{A}^*$, the theory does not require us to do so.

EXAMPLE 2.1. All our examples will have $s \le 4$ and for convenience we will write i, j, k and l instead of i_1, i_2, i_3 and i_4. Whenever no confusion is possible, we will use subscripts and omit the set describing the relevant indices. Thus we will write λ_{ijk} instead of $\lambda_{\{1,2,3\}}(i_1, i_2, i_3)$.

Suppose that $s = 3$ and $\mathscr{A} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$. Then $\Omega_{\mathscr{A}}$ consists of all arrays $f = (f_{ijk})$ representable in the form

$$f_{ijk} = \alpha_{ij} + \beta_{jk} + \gamma_{ki}$$

for some arrays (α_{ij}), (β_{jk}) and (λ_{ki}). □

Of the following interpretations of $M_{\mathscr{A}}$, the first is applicable mainly to a factorial experiment with observations $y = (y(i) : i \in \mathscr{I})$ and expected values $\eta = (\eta(i) : i \in \mathscr{I})$. The others are applicable to a contingency table with cell frequencies $n = (n(i) : i \in \mathscr{I})$ and probabilities $p = (p(i) : i \in \mathscr{I})$ or expected frequencies $\mu = (\mu(i) : i \in \mathscr{I})$.

Additive model: $\eta \in \Omega_{\mathscr{A}}$.
Multiplicative model: $\log p \in \Omega_{\mathscr{A}}$, p positive.
or $\log \mu \in \Omega_{\mathscr{A}}$, μ positive.
Lancaster-additive model I: $p/q \in \Omega_{\mathscr{A}}$.
Lancaster-additive model II: $P/Q \in \Omega_{\mathscr{A}}$.

Here the function $f = \log p$ is defined by $f(i) = \log p(i)$ whilst $f = p/q$ means $f(i) = p(i)/q(i)$ where $q(i) = p_1(i_1) \cdots p_s(i_s)$ is the product of the one-dimensional marginal probabilities from p. Finally $f = P/Q$ means $f(i) = P(i)/Q(i)$ where $P(i) = \sum_{j \le i} p(i)$ and similarly for Q, where $j \le i$ means $j_\sigma \le i_\sigma$, $\sigma = 1, \cdots, s$. Additive and multiplicative models are commonly called *linear* and *log-linear* models, respectively. The general results below apply also to any generalised linear model; see Nelder and Wedderburn (1972), Baker and Nelder (1978).

Why should we study additive, multiplicative and Lancaster-additive models? In the first place, the way in which they combine linearity and economy has an obvious appeal. Less obvious is that they can be characterised by attractive properties relating them to their $\mathscr{A}$-marginal functions; these are given in the following section. Their best-known properties are the no-interaction ones by which they are usually characterised, and these are given in Section 4.

Suppose that f is known or assumed to satisfy $M_{\mathscr{A}}$ so that $f(i)$ is representable as the sum of parameters $\lambda_a(i_a)$. Leaving aside the trivial case when the generating class $\mathscr{A}$ contains only one element, it is always possible to choose more than one parametric representation of f. That is the parameters $\lambda_a(i_a)$ are not uniquely determined by f. The extent to which they are unique is discussed in Section 4.

Generally speaking, the parameters $\lambda_a(i_a)$ have little more than a mathematical existence but, on rare occasions, they also have a physical meaning.

EXAMPLE 2.2. Let i index the cities of a country, let j index age-categories of brides and let k index age-categories of bridegrooms. Let μ_{ijk} be the expected number of marriages, in a given year, in city i between brides of age j and bridegrooms of age k. Then

$$\mu_{ijk} = M_{ij}N_{ik}\rho_{ijk},$$

M_{ij}, N_{ik} being the numbers of eligible women of age j, men of age k in city i at the beginning of the year, and where ρ_{ijk} is the rate of marriages in city i between women of age j and men of age k. It may be very reasonable to assume that $\rho_{ijk} = \rho_{jk}$ so that

$$\log \mu_{ijk} = \log M_{ij} + \log \rho_{jk} + \log N_{ik}.$$

Thus we have an instance of the model of Example 2.1 in which the parameters (α_{ij}), (β_{jk}), (γ_{ik}) can be given a physical interpretation. □

2.2 *Marginals.* For an arbitrary element $\xi = (\xi(i) : i \in \mathscr{I}) \in \Omega$ and a subset $a \subseteq S$ we write $\xi_a(i_a) = \sum_{i_{a'}} \xi(i)$, the sum being over all $i_\sigma = 1, \cdots, r_\sigma$, $\sigma \in a' = S - a$, and call ξ_a the (unweighted) *a-marginal* of ξ. The $\mathscr{A}$-marginals of ξ are $\{\xi_a : a \in \mathscr{A}\}$. Now let $m(i)$ be a positive weight attached to cell i, where $\sum_i m(i) = 1$. It is necessary in much of what follows to work with weight functions different from the uniform weight function $m(i) = (\Pi_\sigma r_\sigma)^{-1}$. We define the ($m$-weighted) *a-marginal mean* $\bar{\eta}_a$ of $\eta \in \Omega$ by

$$\bar{\eta}_a(i_a) = \frac{1}{m_a(i_a)} \sum_{i_{a'}} m(i)\eta(i), \tag{2.2}$$

and the *$\mathscr{A}$-marginal* means of η are $\{\bar{\eta}_a : a \in \mathscr{A}\}$. The ($m$-weighted) *inner product* $\langle \xi, \eta \rangle_m$ of $\xi, \eta \in \Omega$ is defined by

$$\langle \xi, \eta \rangle_m = \sum_i m(i)\xi(i)\eta(i), \tag{2.3}$$

and its associated *norm* (length) is $\| \xi \|_m = \{\langle \xi, \xi \rangle_m\}^{1/2}$.

Additive models. In terms of these notions we can now characterise the additive model $M_{\mathscr{A}}$. We begin with a lemma.

LEMMA 2.1. *Fix ξ, $\eta_0 \in \Omega$ and consider the set of all η with the same $\mathscr{A}$-marginal means as η_0 and the squared distance $\| \eta - \xi \|_m^2$ of each such η from ξ. Suppose that, in this set, there exists η_1 satisfying $\eta_1 - \xi \in \Omega_{\mathscr{A}}$. Then η_1 uniquely minimizes $\| \eta - \xi \|_m^2$.*

PROOF. The condition that η_0 have the same $\mathscr{A}$-marginal means is $\eta - \eta_0 \perp_m \Omega_{\mathscr{A}}$, where orthogonality $\perp_m$ is with respect to the inner product (2.3). Therefore $\eta - \eta_1 \perp_m \Omega_{\mathscr{A}}$ since

$\eta - \eta_0 \perp_m \Omega_{\mathscr{A}}$ and $\eta_1 - \eta_0 \perp_m \Omega_{\mathscr{A}}$. But $\eta_1 - \xi \in \Omega_{\mathscr{A}}$ and so $\langle \eta - \eta_1, \eta_1 - \xi \rangle_m = 0$. Rearrangement gives

$$\| \eta - \xi \|_m^2 - \| \eta_1 - \xi \|_m^2 = \| \eta - \eta_1 \|_m^2$$

which establishes the truth of the Lemma. □

Note that if e is the unit function $e(i) = 1$ and $\xi = ke$, k constant, then $\xi \in \Omega_{\mathscr{A}}$ and it seems appropriate to describe ξ as *uniform*. The characterisation of the additive model can now be stated: any $\eta \in \Omega_{\mathscr{A}}$ is *simplest* in the sense that it is *closest* to being uniform amongst all arrays with the same $\mathscr{A}$-marginal means. Closeness is measured by $\| \cdot \|_m^2$ and simplest means that ξ is uniform. There is thus a separate characterisation for each positive weight function m.

The above discussion has not involved the question of existence, given η_0, ξ, of η_1 satisfying

$$\eta_1 - \eta_0 \perp_m \Omega, \quad \eta_1 - \xi \in \Omega_{\mathscr{A}} \tag{2.4}$$

but this question is well-known to have an affirmative answer. For $\eta_1 - \xi$ is the projection of $\eta_0 - \xi$ onto $\Omega_{\mathscr{A}}$ orthogonal with respect to $\langle \cdot, \cdot \rangle_m$; equivalently, $\eta_0 - \eta_1$ is the orthogonal projection of $\eta_0 - \xi$ onto $\Omega_{\mathscr{A}}^{\perp}$, the orthogonal complement of $\Omega_{\mathscr{A}}$.

Multiplicative models. The analogous characterisation of the multiplicative model $M_{\mathscr{A}}$, which is due to Good (1963) and Ku and Kullback (1968), closely resembles the previous one. Let the (unweighted) $\mathscr{A}$-marginals of the probability p be fixed at those of p_0 and measure the difference between p and a positive probability π by the Kullback discriminatory information

$$K(p, \pi) = \textstyle\sum_i p(i) \log p(i)/\pi(i) = \langle p, \log p/\pi \rangle \tag{2.5}$$

where $\langle \xi, \eta \rangle = \sum_i \xi(i)\eta(i)$ is the unweighted inner product.

LEMMA 2.2. *Suppose that, among all p with the same $\mathscr{A}$-marginals as p_0, there exists p_1 satisfying* $\log p_1/\pi \in \Omega_{\mathscr{A}}$. *Then p_1 uniquely minimises $K(p, \pi)$.*

PROOF. Since $p - p_0 \perp \Omega_{\mathscr{A}}$, $p_1 - p_0 \perp \Omega_{\mathscr{A}}$ and $\log p_1/\pi \in \Omega_{\mathscr{A}}$, we deduce that $\langle p - p_1, \log p_1/\pi \rangle = 0$. Rearranging this gives $\langle p, \log p/\pi \rangle - \langle p_1, \log p_1/\pi \rangle = \langle p, \log p/p_1 \rangle$, i.e. $K(p, \pi) - K(p_1, \pi) = K(p, p_1)$, from which the lemma follows. □

Taking π to be the uniform probability function gives the following characterisation: any p satisfying the multiplicative model $\log p \in \Omega_{\mathscr{A}}$ is simplest in the sense that it maximises $-\sum_i p(i) \log p(i)$ among all probabilities having the same $\mathscr{A}$-marginals. Assuming that $\cup \{a : a \in \mathscr{A}\} = S$, we may take $\pi = q_0$, the product of the one-dimensional marginals of p_0 and obtain the conclusion that any p satisfying the multiplicative model $M_{\mathscr{A}}$ is closest to being independent amongst all probabilities with the same $\mathscr{A}$-marginals, closeness being measured by K.

The existence of p_1 satisfying

$$p_1 - p_0 \perp \Omega_{\mathscr{A}}, \quad \log p_1/\pi \in \Omega_{\mathscr{A}} \tag{2.6}$$

is assured provided that the $\mathscr{A}$-marginals of p_0 admit a positive probability, see Haberman (1974), Barndorff-Nielsen (1978). Darroch and Ratcliff (1972) proved that, with this proviso and for any subspace ω of Ω, it is possible to construct p_1 given p_0, π and ω by generalised iterative scaling. When $\omega = \Omega_{\mathscr{A}}$ then iterative proportional scaling can be used.

Lancaster-additive Model I. The results concerning additive models can be adapted to provide a characterisation of the Lancaster-additive model I and because it is very similar to the two preceding ones, we only give a brief outline.

Suppose that the unweighted $\mathscr{A}$-marginals of p are held fixed at those of p_0, and that

$\cup \{a : a \in \mathscr{A}\} = S$. Then all of the univariate marginals p_σ are also held fixed and so too is $q = \Pi p_\sigma$, equal to q_0 say. If we put $m = q_0$ and $\eta = p/q = p/q_0$ in (2.2) we find that holding p_a fixed is equivalent to holding $\bar{\eta}_a$ fixed. With $\xi = e$ the difference $\| \eta - \xi \|_m^2$ simplifies to

$$\phi^2(p, q_0) = \textstyle\sum_i [p(i) - q_0(i)]^2/q_0(i), \tag{2.7}$$

the Pearson Chi squared measure of difference between p and q_0. Lemma 2.1 may be translated to apply in this context and using that we obtain the following characterisation: any p satisfying the Lancaster-additive model I with $\Omega_\mathscr{A}$ is simplest in the sense of being closest to independent among all probabilities having the same $\mathscr{A}$-marginals, closeness being measured by ϕ^2.

The equations that p_1 satisfies are

$$p_1 - p_0 \perp \Omega_\mathscr{A}, \quad p_1/q_0 \in \Omega_\mathscr{A} \tag{2.8}$$

where $\perp$ here denotes orthogonality with respect to the unweighted inner product. The existence of p_1 given p_0, that is, of a probability function having prescribed $\mathscr{A}$-marginals and satisfying the Lancaster additive model I is not now guaranteed; see Darroch (1974) for a counter-example when $\mathscr{A} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$, and for further comparisons between these models and the analogous multiplicative models.

2.3 *Fitting the models.* Let us suppose that data $y = (y(i) : i \in \mathscr{I})$ from a factorial experiment has a normal distribution with mean $\eta \in \Omega_\mathscr{A}$ and covariance matrix $\sigma^2 \operatorname{diag}(m)^{-1}$, the diagonal matrix with value $m(i)^{-1}$ in the ith position. Then a *sufficient* reduction of y is to the pair $(\eta_1, \| \eta_1 - \eta_0 \|_m^2)$ where η_1, the projection of $y = \eta_0$ onto $\Omega_\mathscr{A}$ orthogonal with respect to $\langle \cdot , \cdot \rangle_m$, satisfies (2.4) with $\xi = 0$. We have already seen that η_1 is completely determined by its $\mathscr{A}$-marginal means, and these coincide with those of y. If we further suppose that m is *completely multiplicative* in that it can be written

$$m(i) = k\, \Pi_\sigma m_\sigma(i_\sigma),$$

where for each $\sigma \in S$, $m_\sigma(i_\sigma) \geq 0$, $\sum_{i_\sigma} m_\sigma(i_\sigma) = 1$ and k is a constant, then we can express η_1 in terms of the $\mathscr{A}$-marginal means of y via formula (3.6) below. Thus (when m is completely multiplicative) the set of $\mathscr{A}$-marginal means is not only a sufficient reduction of y under the *additive model $M_\mathscr{A}$*, but also there is a closed-form solution of the least-squares (= maximum likelihood) estimation problem.

We turn now to the contingency table $n = (n(i) : i \in \mathscr{I})$, supposing that n has a multinomial distribution with probability parameter p satisfying the *multiplicative model $M_\mathscr{A}$* and total sample size $N = \sum_i n(i)$. The (unweighted) $\mathscr{A}$-marginal totals $\{n_a : a \in \mathscr{A}\}$ constitute a sufficient reduction of n and, provided these marginals admit a positive table, the log-likelihood $\langle n, \log p \rangle$ is maximised, or $K((1/N)n, p)$ is minimised, subject to $\log p/\pi \in \Omega_\mathscr{A}$ (normally π is uniform) when $p = p_1$ satisfies (2.6) with $p_0 = (1/N)n$. That these equations give the unique maximum likelihood solution is immediately verified on noting that $\langle \log p_1 - \log p, (1/N)n - p_1 \rangle = 0$ and on rearranging the term on the left-hand side of this equation to give $K((1/N)n, p) - K((1/N)n, p_1) = K(p_1, p)$. As was noted in 2.2 above, the equations (2.6) can be solved by the well-known iterative proportional scaling procedure.

To our knowledge there is no exact maximum-likelihood theory for the fitting of Lancaster-additive multinomial models to contingency tables, although a number of authors have discussed asymptotic theory for likelihood-ratio tests under the independence alternative, see Lancaster (1969) for details.

3. Generalised interactions. Denote by $M_\mathscr{A}$ the model for $f = (f(i) : i \in \mathscr{I})$ defined by

$$M_\mathscr{A} : f \in \Omega_\mathscr{A}.$$

The function f will be variously interpreted as η, $\log p$, p/q or P/Q. In 3.2 below $M_{\mathscr{A}}$ will be formulated as imposing *zero generalised $\mathscr{A}$-interaction,* where generalised interactions are defined very simply by repeatedly averaging over the values $f(i)$ of f.

3.1 *Averaging operators.* Let w_σ be a weight function defined on $\{1, 2, \cdots, r_\sigma\}$, i.e. $\sum_{i_\sigma} w_\sigma(i_\sigma) = 1$. The numbers $w_\sigma(i_\sigma)$ will be thought of as non-negative although there is no strict need for them to be so. Write $S - \{\sigma\} = S - \sigma$. Then the *averaging operator* $T_{S-\sigma}$ operating on f is defined by

$$(T_{S-\sigma}f)(i) = \textstyle\sum_{i_\sigma} w_\sigma(i_\sigma) f(i).$$

Thus $T_{S-\sigma}$ takes weighted averages over the σth coordinate and leaves a function which depends on i through $i_{S-\sigma}$ only. For $a \subseteq S$ let T_a be the operator which takes averages over all coordinates with indices in $a' = S - a$. In other words,

$$T_a = \Pi_{\sigma\in a'} T_{S-\sigma}.$$

For example, if $S = \{1, 2, 3\}$, then $T_{\{1\}} = T_{S-2}T_{S-3} = T_{\{1,3\}}T_{\{1,2\}}$. When $a = S$ we define $T_S = I$, the *identity* operator. An alternative definition of T_a is possible via (2.2): $T_a\eta = \bar{\eta}_a$ where this average is weighted with respect to the completely multiplicative weight function $w(i) = \Pi_{\sigma\in S} w_\sigma(i_\sigma)$. It is immediate that T_a is a linear operator on Ω, that $T_a^2 = T_a$ and, more generally that

$$T_aT_b = T_bT_a = T_{ab} \tag{3.1}$$

where for $a, b \subseteq S$ we write $a \cap b = ab$.

Two particular weight functions w are of special interest. One is the *uniform weight function* defined by

$$w_\sigma(i_\sigma) = 1/r_\sigma.$$

The other is the *substitution weight function* defined by

$$w_\sigma(i_\sigma) = \begin{cases} 0 & \text{if} \quad i_\sigma \neq r_\sigma, \\ 1 & \text{if} \quad i_\sigma = r_\sigma. \end{cases}$$

The resulting substitution operator T_a has the defining property $(T_af)(i) = f(i_ar_{a'})$ where $j = i_ar_{a'}$ denotes the cell with $j_\sigma = i_\sigma$ if $\sigma \in a$ and $j_\sigma = r_\sigma$, if $\sigma \in a'$. Thus T_a substitutes r_σ for i_σ, $\sigma \in a'$. Of course any other fixed reference cell could be used instead of r. It will be convenient to denote $f(i_ar_{a'})$ by $f_a^r(i_a)$.

EXAMPLE 3.1. Let $s = 4$ and $a = \{1, 2\}$. When w is the uniform weight function the transformation $f \to T_af$ replaces f_{ijkl} by $f_{ij\cdot\cdot}$ where, as usual, $\cdot$ denotes uniform average. When w is the substitution weight function f_{ijkl} is replaced under T_a by $f_{ijr_3r_4}$. □

Much of the theory in this paper is obtained using only the simple algebraic equipment of averaging operators. The same ground may be covered using sums and products of linear subspaces and their orthogonal projections. Little will be said about this approach here because it is part of this paper's aim to demonstrate the feasibility of the more elementary approach. It will suffice to show that T_a is an orthogonal projection operator.

We have already noted that $T_a^2 = T_a$ and so T_a is a projection operator. Since $T_af = f$ iff $f(i) = \lambda(i_a)$ it follows that T_a projects onto the subspace Ω_a of Ω defined by this property. Further, T_a is self-adjoint with respect to $\langle\cdot, \cdot\rangle_w$ since

$$\begin{aligned}\langle f, T_ag\rangle_w &= \textstyle\sum_i w(i)f(i)[\sum_{i_{a'}} w_{a'}(i_{a'})g(i)] \\ &= \textstyle\sum_{i_a} w_a(i_a)[\sum_{i_{a'}}(i_{a'})f(i)][\sum_{i_{a'}} w_{a'}(i_{a'})g(i)] = \langle T_af, g\rangle_w.\end{aligned}$$

Finally, T_a is orthogonal with respect to $\langle \cdot , \cdot \rangle_w$ because $\langle (I - T_a)f, T_a f\rangle_w = \langle T_a(I - T_a)f, f\rangle_w = \langle 0, f\rangle_w = 0$.

3.2 *Zero generalised interaction.* Given a generating class $\mathscr{A}$ of subsets of S, define the *generalised $\mathscr{A}$-interaction operator* $I - T_{\mathscr{A}}$ by

$$I - T_{\mathscr{A}} = \Pi_{a\in\mathscr{A}}(I - T_a). \tag{3.2}$$

By (3.1) the terms on the right-hand side of (3.2) can be multiplied together in any order and so, on expanding it, we find

$$T_{\mathscr{A}} = \textstyle\sum_a T_a - \sum\sum_{a\neq b} T_{ab} + \cdots \mp T_{\cap\mathscr{A}} \tag{3.3}$$

where the sums are over all $a \in \mathscr{A}$, distinct pairs $a, b \in \mathscr{A}$, etc. Another useful expression for $T_{\mathscr{A}}$ results from ordering the elements of $\mathscr{A}$ as $a_1, a_2, \cdots, a_m$, namely

$$T_{\mathscr{A}} = T_{a_1} + (I - T_{a_1})T_{a_2} + \cdots + \Pi_{l<m}(I - T_{a_l})T_{a_m}. \tag{3.4}$$

PROPOSITION 3.1. *The function f satisfies $M_{\mathscr{A}}$ if and only if*

$$T_{\mathscr{A}}f = f. \tag{3.5}$$

PROOF. If f satisfies $M_{\mathscr{A}}$ then for some functions $\{\lambda_a : a \in \mathscr{A}\}$ we can write $f = \sum_{a\in\mathscr{A}}\lambda_a$. Now $(I - T_a)\lambda_a = 0$ for each $a \in \mathscr{A}$, and so it follows that $\Pi_{a\in\mathscr{A}}(I - T_a)\sum_{a\in\mathscr{A}}\lambda_a = 0$; that is, $T_{\mathscr{A}}f = f$.

Conversely, if $T_{\mathscr{A}}f = f$ then, by (3.4),

$$f = T_{a_1}f + (I - T_{a_1})T_{a_2}f + \cdots + \Pi_{l<m}(I - T_{a_l})\cdot T_{a_m}f$$

which is of the form $\sum_{a\in\mathscr{A}}\lambda_a$. □

Since the $\{T_a\}$ are orthogonal projections onto the subspaces $\{\Omega_a\}$, it follows that $T_{\mathscr{A}}$ is the orthogonal projection onto $\Omega_{\mathscr{A}} = \sum_{a\in\mathscr{A}}\Omega_a$, although we do not use this fact in what follows.

The proposition formulates $M_{\mathscr{A}}$ as imposing zero generalised $\mathscr{A}$-interaction, in that $(I - T_{\mathscr{A}})f = 0$.

As foreshadowed in Section 2.3 above, when the weight function is completely multiplicative we have an explicit formula for an element satisfying the *additive model* $M_{\mathscr{A}}$ in terms of its $\mathscr{A}$-marginal means, namely

$$\eta = \textstyle\sum_a \bar{\eta}_a - \sum\sum_{a\neq b} \bar{\eta}_{ab} + \cdots \mp \bar{\eta}_{\cap\mathscr{A}}. \tag{3.6}$$

This result is an immediate consequence of (3.3) as soon as we recall that $T_a\eta = \bar{\eta}_a$. Using the substitution weight function we obtain the following special case of (3.6).

$$\eta(i) = \textstyle\sum_a \eta_a^r(i_a) - \sum\sum_{a\neq b} \eta_{ab}^r(i_{ab}) + \cdots \mp \eta_{\cap\mathscr{A}}^r(i_{\cap\mathscr{A}}).$$

From $(I - T_{\mathscr{A}})\log p = 0$ when $T_{\mathscr{A}}$ is based upon the substitution weight function, the multiplicative model is seen to be expressible as

$$\frac{p(i)\cdot \Pi\Pi_{a\neq b}\, p_{ab}^r(i_{ab}) \quad \cdots}{\Pi_a p_a^r(i_a)\cdot \Pi\Pi\Pi_{a\neq b\neq c}\, p_{abc}^r(i_{abc}) \quad \cdots}\,[p_{\cap\mathscr{A}}^r(i_{\cap\mathscr{A}})]^{\pm 1} = 1. \tag{3.7}$$

The left-hand side of (3.7) is a *generalised cross product ratio.*

EXAMPLE 3.2. As in Example 2.1 let $\mathscr{A} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$. Then (3.3) becomes

$$T_{\mathscr{A}} = T_{\{1,2\}} + T_{\{2,3\}} + T_{\{3,1\}} - T_{\{1\}} - T_{\{2\}} - T_{\{3\}} + T_{\phi}.$$

Using the uniform weight function, (3.6) expresses $M_{\mathscr{A}}$ as the familiar

$$\eta_{ijk} = \eta_{ij.} + \eta_{.jk} + \eta_{i.k} - \eta_{i..} - \eta_{.j.} - \eta_{..k} + \eta_{...}$$

while (3.7) becomes the equally familiar cross-product ratio formulation of no three-dimensional interaction, namely

$$\frac{p_{ijk}p_{ir_2r_3}p_{r_1jr_3}p_{r_1r_2k}}{p_{ijr_3}p_{r_1jk}p_{ir_2k}p_{r_1r_2r_3}} = 1. \qquad \square$$

Alternative formulations of the *Lancaster-linear models* $M_{\mathscr{A}}$:$p/q \in \Omega_{\mathscr{A}}$ and $P/Q \in \Omega_{\mathscr{A}}$, will now be given. First choose $w_\sigma = p_\sigma$. Then

$$T_a \frac{p(i)}{q(i)} = \sum_{i_{a'}} q_{a'}(i_{a'}) \frac{p(i)}{q(i)} = \frac{1}{q_a(i_a)} \sum_{i_{a'}} p(i) = \frac{p_a(i_a)}{q_a(i_a)}.$$

Applying (3.3) the Lancaster-additive Model I is seen to be expressible was

$$\frac{p}{q} = \sum_a \frac{p_a}{q_a} - \sum_{a \neq b}\sum \frac{p_{ab}}{q_{ab}} + \cdots \mp \frac{p_{\cap \mathscr{A}}}{q_{\cap \mathscr{A}}}. \tag{3.8}$$

Turning now to the Lancaster-additive model II, let T_a be based on the substitution weight function. Then

$$T_a \frac{P(i)}{Q(i)} = \frac{P(i_a r_{a'})}{Q(i_a r_{a'})} = \frac{P_a(i_a)}{Q_a(i_a)}.$$

Consequently the model here is

$$\frac{P}{Q} = \sum_a \frac{P_a}{Q_a} - \sum_{a \neq b}\sum \frac{P_{ab}}{Q_{ab}} + \cdots \mp \frac{P_{\cap \mathscr{A}}}{Q_{\cap \mathscr{A}}}. \tag{3.9}$$

It is now easy to see that the two Lancaster-additive models are equivalent. After multiplication of (3.8) by $q(i)$ and (3.9) by $Q(i)$, each term in (3.9) is seen to be the distribution function of the corresponding term in (3.8).

3.3 *Marginals and generalised interactions.* A by-product of the model characterisations of 2.2 above is that, given $f \in \Omega_{\mathscr{A}}$, where f is η, $\log p$, p/q or P/Q, f is uniquely determined by its $\mathscr{A}$-marginals, suitably interpreted as weighted means or unweighted sums. This is a special case of the result which we now prove that given its $\mathscr{A}$-marginals and its generalised $\mathscr{A}$-interaction, f is uniquely determined.

There is almost nothing in the proof for η, p/q, P/Q. Thus, defining $T_{\mathscr{A}}$ with respect to any completely multiplicative weight function w, we can write $\eta = T_{\mathscr{A}}\eta + (I - T_{\mathscr{A}})\,\eta$ as the sum of the expansion (3.6), involving its $\mathscr{A}$-marginals, and its generalised $\mathscr{A}$-interaction. Similarly for p/q, except that we now define $T_{\mathscr{A}}$ with respect to $w = q$ and use (3.8), and for P/Q where the substitution operators are used.

There is no explicit demonstration of this uniqueness result for $\log p$ and it has to be proved using Lemma 2.2. Let us suppose that p is a positive probability and that $(I - T_{\mathscr{A}})\log p = u$. Define $\pi = k \exp u$ where k is the normalising constant making $\sum_i \pi(i) = 1$. Then $T_{\mathscr{A}}\log p/\pi = T_{\mathscr{A}}(\log p - \log k - u) = T_{\mathscr{A}}\log p - \log k = \log p - \log k - u$ by the definition of u and the fact that $T_{\mathscr{A}}u = 0$. But this means that $\log p/\pi \in \Omega_{\mathscr{A}}$ and by Lemma 2.2 there is only one p with this property having given $\mathscr{A}$-marginal sums, provided only that these marginals admit a positive probability.

A postscript on this result is the following: it does not matter which (completely multiplicative) weight function w is used to define the generalised $\mathscr{A}$-interaction function $(I - T_{\mathscr{A}})f$ because $(I - T_{\mathscr{A}})f$ defined with respect to one weight function is recoverable from $(I - T_{\mathscr{A}})f$ defined with respect to another. For, if $\{T_a\}$ and $\{\tilde{T}_a\}$ are defined with respect to w and $\tilde{w}$, we see from $\tilde{T}_{\mathscr{A}}T_a = T_a$, $a \in \mathscr{A}$, and (3.3) that $\tilde{T}_{\mathscr{A}}T_{\mathscr{A}} = T_{\mathscr{A}}$, i.e. that

$$(I - \tilde{T}_{\mathscr{A}})(I - T_{\mathscr{A}})f = (I - \tilde{T}_{\mathscr{A}})f.$$

Incidentally, this identity shows directly why $T_{\mathscr{A}}f = f$ iff $\tilde{T}_{\mathscr{A}}f = f$, a fact implicit in Proposition 3.1.

4. Interactions.

4.1 *Interaction operators.* In the previous section we saw that, given a weight function w and averaging operators T_a, the operators $T_{\mathscr{A}}$ and $I - T_{\mathscr{A}}$ arise naturally from consideration of the model $M_{\mathscr{A}}$. In the particular case $\mathscr{A} = \{S - \sigma: \sigma \in S\}$ the operator $I - T_{\mathscr{A}} = \Pi_{\sigma \in S}(I - T_{S-\sigma})$ will be denoted by U_S and called the S-interaction operator. Thus

$$U_S = \Pi_{\sigma \in S}(T_S - T_{S-\sigma}).$$

The definition is now extended to cover any subset b of S. Define $U_\phi = T_\phi$ and, otherwise

$$U_b = \Pi_{\sigma \in b}(T_b - T_{b-\sigma}).$$

The operator U_b will be called the b-interaction operator. Alternative ways of writing it are easily seen to be

$$U_b = \Pi_{\sigma \in b}(I - T_{b-\sigma}).\, T_b, \tag{4.1}$$

$$U_b = \Pi_{\sigma \in b}(I - T_{S-\sigma}).\, \Pi_{\sigma \in b'} T_{S-\sigma}, \tag{4.2}$$

$$U_b = \textstyle\sum_{c \subseteq b} (-1)^{|b-c|} T_c. \tag{4.3}$$

EXAMPLE 4.1. Again let $s = 3$. The interaction operator $U_{\{1,2,3\}}$ is identical to the operator $I - T_{\mathscr{A}}$, with $\mathscr{A} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$, discussed in Example 3.2. The interaction operator $U_{\{1,2\}}$ is expressible in various ways as

$$U_{\{1,2\}} = (I - T_{\{2\}})(I - T_{\{1\}})T_{\{1,2\}} = (I - T_{\{2,3\}})(I - T_{\{1,3\}})T_{\{1,2\}}$$
$$= T_{\{1,2\}} - T_{\{1\}} - T_{\{2\}} + T_\phi.$$

Thus, for the uniform weight function,

$$(U_{\{1,2\}}\eta)_{ijk} = \eta_{ij\cdot} - \eta_{i\cdot\cdot} - \eta_{\cdot j \cdot} + \eta_{\cdot\cdot\cdot}. \quad \square$$

Interactions are usually introduced recursively and their recursive structure is clearly seen in the interaction operators. For example, when $s = 3$,

$$U_{\{1,2,3\}} = (I - T_{\{2,3\}})(I - T_{\{1,3\}}) - (I - T_{\{2,3\}})(I - T_{\{1,3\}})T_{\{1,2\}}.$$

The second term on the right side is $U_{\{1,2\}}$ and gives $\{1, 2\}$ interactions averaged over k. The first term gives $\{1, 2\}$ interactions within each level k. Thus $\{1, 2, 3\}$ interactions are clearly seen to be differences of $\{1, 2\}$ interactions.

Some basic results about interaction operators are collected together in the following lemma.

LEMMA 4.1 (i) $T_a U_b = 0$ *if* $b \not\subseteq a$.

(ii) $\sum_{i_\sigma} w_\sigma(i_\sigma) U_b f(i) = 0$ *if* $\sigma \in b$.

(iii) $T_a U_b = U_b$ *if* $b \subseteq a$.

(iv) $\sum_{b \subseteq a} U_b = T_a$.

(v) $U_b^2 = U_b$.

(vi) $U_a U_b = 0$ *if* $a \neq b$.

(vii) *Let* $b_1, \cdots, b_m$ *be distinct sets. Then*

$\sum_j k_j U_{b_j} f = 0$ *implies that* $k_j U_{b_j} f = 0$ *for all* j.

(viii) U_b *is self-adjoint with respect to the inner product* $\langle \cdot, \cdot \rangle_w$.

PROOF (i) Choose $\tau \in b - a$. Since $T_a = \Pi_{\sigma\in a'} T_{S-\sigma}$, it follows that $T_{S-\tau}(I - T_{S-\tau})$ is a factor of $T_a U_b$.

(ii) By (i) $T_{S-\sigma} U_b = 0$, $\sigma \in b$.

(iii) Apply (4.1) and (3.1).

(iv) First consider $a = S$. By (4.2) $\sum_{b\subseteq S} U_b = \sum_{b\subseteq S} [\Pi_{\sigma\in b}(I - T_{S-\sigma}) \Pi_{\sigma\in b'} T_{S-\sigma}] = \Pi_{\sigma\in S}[(I - T_{S-\sigma}) + T_{S-\sigma}] = \Pi_{\sigma\in S} I = I$. Having established that $\sum_{b\subseteq S} U_b = I$, we now multiply by T_a to get $\sum_{b\subseteq S} T_a U_b = T_a$. Application of (i) and (iii) now gives (iv).

(v) U_b is a product of idempotent operators which commute and hence is itself idempotent.

(vi) Choose $\tau \in (b - a) \cup (a - b)$ and reason as in the proof of (i).

(vii) Multiply $\sum_j k_j U_{b_j} f = 0$ by U_b and apply (v) and (vi).

(viii) By (4.3) U_b is a linear combination of operators which are self-adjoint. □

We note that U_b is an orthogonal projection operator because it is idempotent and self-adjoint. Further $U_b f(i) = g(i_b)$ say and for each $\sigma \in b$, $\sum_{i\sigma} w_\sigma(i_\sigma) g(i_b) = 0$. Moreover, if f is a function satisfying (a) $f(i) = h(i_b)$ and (b) $\sum_{i_\sigma} w_\sigma(i_\sigma) f(i) = 0$ for all $\sigma \in b$, then, by (4.1), $U_b f = f$. Thus U_b is the orthogonal projection operator onto the subspace Θ_b of all functions satisfying (a) and (b), although we will not use this interpretation in the sequel.

4.2 *Hierarchical no-interaction models.* Let the *closure* $\bar{\mathscr{A}}$ of a generating class $\mathscr{A}$ be defined by

$$\bar{\mathscr{A}} = \{b: b \subseteq a \quad \text{for some} \quad a \in \mathscr{A}\}.$$

The complement of $\bar{\mathscr{A}}$ is

$$\bar{\mathscr{A}}' = \{b: b \not\subseteq a \quad \text{for all} \quad a \in \mathscr{A}\}.$$

Note that the class $\bar{\mathscr{A}}'$ is hierarchical. That is, if $b_1 \in \bar{\mathscr{A}}'$ and $b_2 \supseteq b_1$, then $b_2 \in \bar{\mathscr{A}}'$.

PROPOSITION 4.1. $T_{\mathscr{A}} = \sum_{b\in\bar{\mathscr{A}}} U_b$.

PROOF. It is easier to prove that

$$I - T_{\mathscr{A}} = \sum_{b\in\bar{\mathscr{A}}'} U_b \tag{4.4}$$

from which the proposition follows. But this is a direct consequence of our definitions and Lemma 4.1. For

$$I - T_{\mathscr{A}} = \Pi_{a\in\mathscr{A}}(I - T_a) \text{ by the definition (3.2)}$$

$$= \Pi_{a\in\mathscr{A}}(\textstyle\sum_{b\not\subseteq a} U_b) \text{ by (iv) of Lemma 4.1}$$

$$= \textstyle\sum_{b\in\bar{\mathscr{A}}'} U_b \text{ by (v) and (vi) of Lemma 4.1,}$$

and the definition of $\bar{\mathscr{A}}'$. □

Thus the model $M_{\mathscr{A}}$ for f may now be expressed as

$$f(i) = \sum_{b\in\bar{\mathscr{A}}} U_b f(i) \tag{4.5}$$

or as

$$U_b f(i) = 0 \quad \text{for all} \quad b \in \bar{\mathscr{A}}'. \tag{4.6}$$

Formula (4.5) follows immediately from Proposition 4.1 and formula (4.6) by application of (iv) with $a = S$ and (vii) of Lemma 4.1. By virtue of (4.6), $M_{\mathscr{A}}$ may be called a *hierarchical no-interaction model.* Proposition 4.1 thus provides the link with the more common approach to models and interactions which starts with interactions and then defines models by requiring that a hierarchical set of interactions are zero.

Models with equal sized generating sets are frequently used in searches for parsimonious fits to data and, for such models, there is a simple formula relating $T_{\mathscr{A}}$ to $\{T_b: b \in \bar{\mathscr{A}}\}$.

EXAMPLE 4.2. Let $s = 5$ and consider $\mathscr{A} = \{12, 13, 14, 15, 23, 24, 25, 34, 35, 45\}$ where 12 denotes $\{1, 2\}$ etc. We shall prove that

$$T_{\mathscr{A}} = [T_{12} + \cdots + T_{45}] - 3[T_1 + \cdots + T_5] + 6T_\phi. \quad \square$$

The general result is given in Proposition 4.2 below. It really belongs in Section 3 but its proof uses results of this section.

PROPOSITION 4.2. *For* $0 \le t < s$ *let*

$$\mathscr{A} = \mathscr{A}_t^s = \{a \subseteq S: |a| = t\}.$$

Then

$$T_{\mathscr{A}} = \sum_{u=0}^{t} (-1)^{t-u}\binom{s-u-1}{t-u} \sum_{b:|b|=u} T_b.$$

PROOF.

$$\begin{aligned}
T_{\mathscr{A}} &= \sum_{b:|b|\le t} U_b = \sum_{b:|b|\le t} \Pi_{\sigma\in b}(I - T_{S-\sigma})\,\Pi_{\sigma\in b'} T_{S-\sigma} \\
&= \sum_{u=0}^{t} \text{coefficient of } z^u \text{ in } \Pi_{\sigma\in S}[z(I - T_{S-\sigma}) + T_{S-\sigma}] \\
&= \text{coefficient of } z^t \text{ in } (1-z)^{-1}\,\Pi_{\sigma\in S}[zI + (1-z)T_{S-\sigma}] \\
&= \text{coefficient of } z^t \text{ in } (1-z)^{-1} \sum_{u=0}^{s} z^u(1-z)^{s-u} \sum_{b:|b|=u} T_b \\
&= \text{coefficient of } z^t \text{ in } \sum_{u=0}^{t} z^u(1-z)^{s-u-1} \sum_{b:|b|=u} T_b \\
&= \sum_{u=0}^{t} (-1)^{t-u}\binom{s-u-1}{t-u} \sum_{b:|b|=u} T_b. \quad \square
\end{aligned}$$

4.3 *Dimensions of models.* Let us denote the rank of a linear operator T by $r(T)$, and the dimension of a subspace ω of Ω by $\dim \omega$. The following are immediate consequences of the relevant definitions.

$$\dim \Omega_a = r(T_a) = \Pi_{\sigma\in a} r_\sigma.$$

$$\dim \Theta_b = r(U_b) = \Pi_{\sigma\in b}(r_\sigma - 1).$$

Our next result is an immediate consequence of Propositions 3.1, 4.1, and the linearity of trace, as soon as we recall that $r(P) = \text{trace}\,(P)$ for a projection operator P.

PROPOSITION 4.3. (i) *For a generating class* $\mathscr{A}$

$$\dim \Omega_{\mathscr{A}} = \sum_a \Pi_{\sigma\in a} r_\sigma - \sum_{a\ne b}\sum \Pi_{\sigma\in ab} r_\sigma + \cdots \mp \Pi_{\sigma\in\cap\mathscr{A}} r_\sigma = \sum_{b\in\bar{\mathscr{A}}} \Pi_{\sigma\in b}(r_\sigma - 1).$$

(ii) *For any t satisfying* $0 \le t < |S| = s$

$$\dim \Omega_{\mathscr{A}_t^s} = \sum_{u=0}^{t} (-1)^{t-u}\binom{s-u-1}{t-u} \sum_{b:|b|=u} \Pi_{\sigma\in b} r_\sigma.$$

4.4 *Discussion.* As an illustration of the use of the interaction operators with the *additive model* $M_{\mathscr{A}}$, consider the following simple method of deriving the least-squares estimates of the interactions $U_b\eta(i)$ of η when we have data $y = (y(i,j): j = 1, \cdots, n(i), i \in \mathscr{I})$ with $n(i)$ observations made on cell i, and the cell frequencies $n(i)$ are proportional, that is, completely multiplicative

$$n(i) = \Pi_\sigma n_\sigma(i_\sigma)/N^{s-1} \tag{4.7}$$

where $N = \sum_i n(i)$. Condition (4.7) is of course most likely to be realised in practice when $n(i)$ is constant. If we denote the mean of $y(i, j)$ over j by $y(i)$ then the sum of squared deviations of the observations from their expectations is $\sum_{i,j} (y(i, j) - \eta(i))^2 = \sum_{i,j} (y(i, j) - y(i))^2 + \sum_i n(i)(y(i) - \eta(i))^2$, so that the least value has to be found of

(4.8) $$\sum_i n(i)(y(i) - \eta(i))^2 = \sum_{b \subseteq S} \sum_{i_b} n_b(i_b)(U_b y(i) - U_b \eta(i))^2.$$

Identity (4.8) follows from the calculation

$$\langle z, z \rangle_w = \langle z, \sum_b U_b z \rangle_w = \sum_b \langle z, U_b^2 z \rangle_w = \sum_b \langle U_b z, U_b z \rangle_w,$$

using Lemma 4.1 (iv), (v) and (viii), with $z = y - \eta$ and $w(i) = n(i)/N$. Identity (4.8) shows that, for any no-interaction model (hierarchical or not), the least squares estimate of $U_b \eta$ is $U_b y$ (when the cell frequencies are proportional) for every model in which this interaction is not assumed zero.

Consider now the *multiplicative model* $M_{\mathscr{A}}$, i.e. $\log p \in \Omega_{\mathscr{A}}$. Two particular weight functions have been widely used in the literature. Since Birch's (1963) paper, most authors have used the uniform weight function. In this case

$$U_b \log p(i) = \sum_{c \subseteq b} (-1)^{|b-c|} \log p_c^*(i_c)$$

where $p_c^*(i_c)$ is the *geometric mean* of all $p(j)$ for which $j_c = i_c$, and we do not find these interactions easy to interpret. The system of interactions based upon the substitution weight function does seem easier to interpret with multiplicative models and has been used to effect by Plackett (1974). It was introduced by Mantel (1966), and is used more generally in GLIM, see Baker and Nelder (1978). Here

$$U_b \log p(i) = \sum_{c \subseteq b} (-1)^{|b-c|} \log p_c^r(i_c),$$

which is the logarithm of a cross product ratio. Thus if $d = 3$ and $b = \{1, 2\}$, the cross product ratio is

$$\frac{p(i, j, r_3)p(r_1, r_2, r_3)}{p(i, r_2, r_3)p(r_1, j, r_3)}.$$

Referring back to Section 2.3 above, we now turn to what may be called the estimated model interactions $U_b \log \hat{p}$, $b \in \bar{\mathscr{A}}$, where $\hat{p}$ is the maximum likelihood estimate of p under $M_{\mathscr{A}}$. No matter which w is chosen, $U_b \log \hat{p}$ does not share the attractive properties of $U_b \hat{\eta}$ when the cell frequencies are proportional, properties which stem from the equation $U_b \hat{\eta} = U_b y$. Thus $U_b \log \hat{p}$ does not depend only on the b-marginal table n_b of $n = (n(i): i \in \mathscr{I})$ but, in general, on all $\mathscr{A}$-marginals. (An important exception occurs when the generating class is decomposable; see Haberman (1974), Darroch, Lauritzen and Speed (1980) and Lauritzen, Speed and Vijayan (1978).) Also it changes each time a different model (that is, a different $\mathscr{A}$) is fitted. This is one of the most important differences between the additive and multiplicative models. Of course when b is one of the maximal elements of $\mathscr{A}$, that is $b \in \mathscr{A}^*$, then $U_b \log \hat{p}$ can be put to use since its magnitude, relative to its standard deviation, indicates whether or not the model obtained from $M_{\mathscr{A}}$ by putting $U_b \log p = 0$ is likely to be acceptable; see Baker and Nelder (1978).

Finally we consider the implications of Proposition 4.1 for *Lancaster-additive models.* Using the weight function q (see Section 3.2) the b-interaction for model I is

(4.9) $$U_b \frac{p}{q} = \sum_{c \subseteq b} (-1)^{|b-c|} \frac{p_c}{q_c},$$

and using the substitution weight function the b-interaction for model II is

(4.10) $$U_b \frac{P}{Q} = \sum_{c \subseteq b} (-1)^{|b-c|} \frac{P_c}{Q_c}.$$

It is easy to see (cf. Section 3.2) that the two definitions of no b-interaction obtained from

(4.9) and (4.10) are equivalent to each other and to Lancaster's (1969, page 256) definition, namely

(4.11) $$\Pi_{\sigma\in b}(P_\sigma^*(i_\sigma) - P_\sigma(i_\sigma)) = 0,$$

where the P_σ^* are artificial functions multiplied according to the rule

$$\Pi_{\sigma\in c}P_\sigma^*(i_\sigma) = P_c(i_c).$$

Zentgraf (1975) proved that, if (4.11) holds for all b with $|b| > t$, then

(4.12) $$P(i) = \sum_{u=0}^{t} (-1)^{t-u}\binom{s-u-1}{t-u} \sum_{b:|b|=u} P_b(i_b)Q_{b'}(i_{b'}).$$

This result, when combined with its converse, amounts to a special case of Proposition 4.2 above.

4.5 *A uniqueness property of interactions.* The main purpose of this paper has been to show that many general properties linking models and interactions can be easily stated and proved using interaction operators. We have seen that given any model $M_{\mathscr{A}}$ and any multiplicative weight function w there corresponds a generalized interaction operator $T_{\mathscr{A}}$, that the interaction operators U_b provide a useful way of partitioning $T_{\mathscr{A}}$ and, finally, that $M_{\mathscr{A}}$ has the "hierarchical no-interaction" property by which it is usually characterised.

We conclude by returning to a question raised in Section 2.1, namely: given that f satisfies $M_{\mathscr{A}}$, to what extent are the parameters $\lambda_a(i_a)$ uniquely determined by f? The answer, as shown in the following proposition, is that interactions and only interactions of λ_a are uniquely determined.

PROPOSITION 4.4. *Assume*

(4.13) $$f(i) = \sum_{a\in\mathscr{A}}\lambda_a(i_a)$$

and let $c \in \mathscr{A}$. The extent to which λ_c is determined by f is defined by the equations

(4.14) $$U_b\lambda_c(i_c) = U_bf(i) \quad \text{for all} \quad b \in \bar{\mathscr{A}} - \bar{\mathscr{C}}$$

where $\mathscr{C} = \mathscr{A} - \{c\}$ and where the U_b are defined with respect to any multiplicative weight function.

PROOF. Since

$$f(i) - \lambda_c(i_c) = \sum_{a\in\mathscr{C}}\lambda_a(i_a)$$

therefore

$$U_b(f(i) - \lambda_c(i_c)) = 0 \quad \text{for all} \quad b \in \bar{\mathscr{C}}'.$$

However $U_bf(i) = U_b\lambda_c(i_c) = 0$ for all $b \in \bar{\mathscr{A}}'$. Thus, given (4.13), λ_c certainly satisfies (4.14).

We now prove that equations (4.14) define *all* that is uniquely determined about λ_c from a knowledge of f. This is done by showing that the information about λ_c contained in (4.14) is sufficient for us to construct a λ_c, λ_c^* say, such that

$$f(i) = \lambda_c^*(i_c) + \sum_{a\in\mathscr{C}}\lambda_a(i_a).$$

Simply define

$$\lambda_c^* = \sum_{b\in\bar{\mathscr{A}}-\bar{\mathscr{C}}} U_bf.$$

Then

$$f(i) - \lambda_c^*(i_c) = \sum_{b\in\bar{\mathscr{C}}} U_bf(i)$$

and, by Proposition 4.1, the right side can be written in the form

$$\sum_{a\in\mathscr{C}}\lambda_a(i_a). \quad \square$$

EXAMPLE 2.2 (continued). We have $s = 3$ and $\mathscr{A} = \{\{1, 2\}, \{2, 3\}, \{3, 1\}\}$. Let $c = \{2, 3\}$ so that $\bar{\mathscr{A}} - \bar{\mathscr{C}} = \{\{2, 3\}\}$. Using the substitution weight function for convenience, we find that the total information about the marriage rates ρ_{jk} that can be determined from a knowledge of the expected numbers of marriages μ_{ijk} is contained in the equations

$$\frac{\rho_{jk}\rho_{r_2r_3}}{\rho_{jr_3}\rho_{r_2k}} = \frac{\mu_{r_1jk}\mu_{r_1r_2r_3}}{\mu_{r_1jr_3}\mu_{r_1r_2k}}.$$

Likewise, all that can be determined about the numbers M_{ij} of eligible women is contained in the equations

$$\frac{M_{ij}M_{r_1r_2}}{M_{ir_2}M_{r_1j}} = \frac{\mu_{ijr_3}\mu_{r_1r_2r_3}}{\mu_{ir_2r_3}\mu_{r_1jr_3}}. \quad \square$$

Acknowledgement. We are grateful to the referees and an Associate Editor for their helpful comments.

REFERENCES

ANDERSEN, A. H. (1974). Multidimensional contingency tables. *Scand. J. Statist.* **1** 115–127.
BAKER, R. J. and NELDER, J. A. (1978). *The GLIM System, Release* 3. *Generalised Linear Interactive Modelling.* The Numerical Algorithms Group, Oxford.
BARNDORFF-NIELSEN, O. (1978). *Information and Exponential Families.* Wiley, Chichester.
BIRCH, M. W. (1963). Maximum likelihood in three-way contingency tables. *J. Roy. Statist. Soc. Ser. B* **25** 220–233.
BISHOP, Y. M. M., FIENBERG, S. E. and HOLLAND, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice.* M.I.T. Press, Cambridge, Mass.
DARROCH, J. N. (1974). Multiplicative and additive interactions in contingency tables. *Biometrika* **61** 207–214.
DARROCH, J. N., LAURITZEN, S. L. and SPEED, T. P. (1980). Markov fields and log-linear interaction models for contingency tables. *Ann. Statist.* **8** 522–539.
DARROCH, J. N. and RATCLIFF, D. (1972). Generalized iterative scaling for log-linear models. *Ann. Math. Statist.* **43** 1470–1480.
DAVIDSON, R. (1973). Determination of confounding. *Stochastic Analysis.* Edited by D. G. Kendall and E. P. Harding. Wiley, New York.
FIENBERG, S. E. (1977). *The Analysis of Cross-Classified Data.* M.I.T. Press, Cambridge, Mass.
GOKHALE, D. V. and KULLBACK, S. (1978). *The Information in Contingency Tables.* Dekker, New York.
GOOD, I. J. (1958). The interaction algorithm and practical Fourier analysis. *J. Roy. Statist. Soc. Ser. B* **20** 361–372.
GOOD, I. J. (1963). Maximum entropy for hypothesis formulation, especially for multidimensional contengency tables. *Ann. Math. Statist.* **34** 911–934.
GOODMAN, L. A. (1970). The multivariate analysis of qualitative data: interaction among multiple classifications. *J. Amer. Statist. Assoc.* **65** 226–256.
GRIZZLE, J. E., STARMER, C. F. and KOCH, G. G. (1969). Analysis of categorical data by linear models. *Biometrics* **25** 489–504.
HABERMAN, S. J. (1974). *The Analysis of Frequency Data.* Univ. of Chicago Press, Chicago, Il.
HABERMAN, S. J. (1975). Direct products and linear models for complete factorial tables. *Ann. Statist.* **3** 314–333.
HABERMAN, S. J. (1978). *Analysis of Qualitative data,* Vol. 1. Academic, New York.
HABERMAN, S. J. (1979). *Analysis of Qualitative data,* Vol. 2. Academic, New York.
KU, H. H. and KULLBACK, S. (1968). Interaction in multidimensional contingency tables: an information theoretic approach. *J. Res. Nat. Bur. of Stand.* **72B** 159–199.
KURKJIAN, B., and ZELEN, M. (1962). A calculus for factorial arrangements. *Ann. Math. Statist.* **33** 600–619.
LANCASTER, H. O. (1969). *The Chi-Squared Distribution.* Wiley, New York.
LAURITZEN, S. L., SPEED, T. P. and VIJAYAN, K. (1978). *Decomposable Graphs and Hypergraphs.* Preprint No. 9. Institute of Mathematical Statistics, University of Copenhagen.

MANN, H. B. (1949). *Analysis and Design of Experiments.* Dover, New York.

MANTEL, N. (1966). Models for complex contingency tables and polychotomous dosage response curves. *Biometrics* **22** 83–95.

NELDER, J. A. and WEDDERBURN, R. W. M. (1972). Generalised linear models. *J. Roy. Statist. Soc. Ser.* A **135** 370–384.

PLACKETT, R. L. (1981). *Analysis of Categorical Data*, 2nd Ed. Griffin, London.

ZENTGRAF, R. (1975). A note on Lancaster's definition of higher-order interactions. *Biometrika* **62** 375–378.

SCHOOL OF MATHEMATICAL SCIENCES,
THE FLINDERS UNIVERSITY OF SOUTH AUSTRALIA,
ADELAIDE. S.A. 5042,
AUSTRALIA.

D.M.S., C.S.I.R.O.,
P.O. Box 1965,
CANBERRA. A.C.T. 2601,
AUSTRALIA.