Quasi-symmetry and representation theory

by

Peter McCullagh
Department of Statistics
University of Chicago
July 2001

Summary

Quasi-symmetry is a model for square contingency tables with rows and columns indexed by the same set. In a log-linear model, quasi-symmetry is associated with a vector subspace QS_n of certain real-valued functions on $\mathbf{n} \times \mathbf{n}$. Permutation of factor levels induces a linear transformation $QS_n \to QS_n$, making this into a representation of the symmetric group S_n. However, there are many group representations that are totally unsatisfactory as linear or generalized linear models. Quasi-symmetry is also a hereditary group representation, which is to say that the restriction of QS_{n+1} to the leading $n \times n$ sub-array is equal to QS_n. Another way of saying the same thing is that, as a sequence of vector spaces, QS is a representation of the category of injective maps on finite sets. This property is of fundamental importance for statistical models. This paper sets out to list all hereditary sub-representations by real-valued square matrices, and to explain how these may used in model construction. We conclude that there are exactly six non-trivial log-linear models for a square contingency table.

1. Introduction

A real-valued square matrix f is said to be quasi-symmetric of order n, if there exists a vector $g \in \mathcal{R}^n$ and a symmetric matrix h such that

$$f_{ij} = g_i + h_{ij}. \tag{1}$$

Equivalently, since $h'_{ij} = g_i + g_j + h_{ij}$ is symmetric, the condition can be written in the form $f_{ij} = h_{ij} + g_j$. The original definition of quasi-symmetry, given by Caussinus (1965) in the context of positive matrices and log-linear models, involves multiplication rather than addition. That is to say, a positive matrix m is quasi-symmetric in the original sense if the matrix $\log m$ with components $\log m_{ij}$ is quasi-symmetric in the sense of (1).

If f, f' are both quasi-symmetric, there exists a vector g' and a symmetric matrix h' such that $f'_{ij} = g'_i + h'_{ij}$. Consequently, for all scalars α, α', the linear combination $\alpha f + \alpha' f'$ is also quasi-symmetric. That is to say, the set of quasi-symmetric matrices of order n is a vector space. We write $QS_n \subset \mathcal{R}^{n^2}$, noting that, for $n \geq 1$, QS_n has dimension $(n^2 + 3n - 2)/2$. The co-dimension for the residual degrees of freedom is $(n-1)(n-2)/2$.

As many applied statisticians are well aware, quasi-symmetry is no ordinary vector subspace. It arises as a statistical model in all manner of apparently unrelated applications, such as the following.

(i) Case-control studies in epidemiology (Pike, Casagrande and Smith 1975; Darroch 1981; McCullagh 1982);

(ii) Reversible Markov chains (Kelly, 1979);
(iii) Gravity models in migration studies, (Stewart, 1948; Upton, 1985);
(iv) Transmission disequilibrium models in genetics, (Spielman et al. 1993; Ewens and Spielman 1995; and Sethuraman and Speed 1997),
(v) Matched pairs designs (Fienberg and Larntz 1976),
(vi) Citation studies in science (Stigler, 1994).

The closely related Bradley-Terry model (Bradley and Terry, 1952) is frequently used for assessing team strengths (Barry and Hartigan 1993), and for rating teams or players in tournaments (Joe 1990). Although the derivations seem to have little in common, it is hard to believe that there is not a common theme.

The chief aim of this paper is to uncover what it is that distinguishes quasi-symmetry from ordinary run-of-the-mill vector subspaces, and to describe the entire class of vector subspaces that have the same properties. Quasi-symmetry has two properties, both relevant to statistical models, that set it apart from the great majority of subspaces. The first property is connected with representation theory for the symmetric group (Diaconis, 1988), the idea here being that the order in which the factor levels are listed is sometimes entirely arbitrary. A permutation $\pi \in S_n$ acts naturally on vectors $f \in \mathcal{R}^n$ by composition such that f is carried to $\pi^* f$ with components $(\pi^* f)_i = f_{\pi(i)}$. The same permutation also acts naturally on square matrices in $\mathcal{R}^{n^2}$ by composition such that the matrix f is carried to $\pi^* f$ by

$$(\pi^* f)_{ij} = f_{\pi_i, \pi_j}.$$

This is a linear transformation $\mathcal{R}^{n^2} \to \mathcal{R}^{n^2}$, which could be expressed in the form of matrix multiplication if that were helpful. If f is quasi-symmetric with the decomposition (1), the image $\pi^* f$ is also quasi-symmetric:

$$(\pi^* f)_{ij} = g_{\pi_i} + h_{\pi_i \pi_j} = g'_i + h'_{ij}$$

such that h' is symmetric. In algebraic terminology, the association of each group element π with a linear transformation $\pi^*\colon QS_n \to QS_n$ is a representation $*\colon S_n \to GL(QS_n)$ of the symmetric group by invertible linear transformations $QS_n \to QS_n$ on a vector space. The induced linear transformation is not in fact a homomorphism in the usual sense, but an anti-homomorphism: the group product $\pi\varphi$ is carried to the composition $\varphi^*\pi^*\colon \mathcal{R}^{n^2} \to \mathcal{R}^{n^2}$ in reverse order. This reversal is of little consequence for group representations, but it is important in the critical property of inheritance.

The second property of quasi-symmetry is not so much a property of the vector subspace $QS_n \subset \mathcal{R}^{n^2}$ as a property of the *sequence* of subspaces $\{QS_1 \subset \mathcal{R}^{1^2}, QS_2 \subset \mathcal{R}^{2^2}, \ldots\}$. That is to say, the second property applies to quasi-symmetry as a statistical model formula, defined in a similar manner for each integer $n \geq 0$. If $f \in QS_n$, the leading $(n-1) \times (n-1)$ sub-matrix is also quasi-symmetric, and thus in QS_{n-1}. This humble inheritance property guarantees that the sub-models $QS_4 \subset \mathcal{R}^{4^2}$, $QS_5 \subset \mathcal{R}^{5^2}$ and so on, are alike and have similar interpretations. Generally speaking, the number of recorded levels is rather arbitrary, and determined to some extent by the investigator through selection and aggregation. Any pattern that arises as a consequence of such capricious choices is of little scientific interest. For a model to be sensible, it must be immune, or at least robust, to selection of levels and, if relevant, aggregation of levels. Selection and aggregation are not invertible transformations, so group representations do not begin to address this inheritance property, which is the

critical ingredient that distinguishes a possibly sensible statistical model from a definitely silly model (McCullagh, 2000).

It is easy to see, for example that the diagonal matrices, the symmetric matrices and the skew-symmetric matrices are group-invariant and also have the inheritance property. So, although quasi-symmetry is very special, it is certainly not unique in this respect. With a little effort, it is possible to exhibit a number of other group-invariant sequences, such as the constant matrices, that also have the inheritance property. The purpose of this paper is to provide a complete catalogue of statistical models that occur as sub-representations in $\mathcal{R}^{n^2}$.

2. Representation of injective maps

2.1 *Model formula*

We set out in this section to identify all model formulae for square two-way tables with rows and columns indexed by the same labels. For this purpose, a model formula is defined as follows. For each $n \geq 0$ a model formula M identifies a vector subspace $M_n \subset \mathcal{R}^{n\times n}$ that is closed under simultaneous permutation of rows and columns. Moreover, the sequence $\{M_n\}$ is such that M_n is embedded in M_{n+1} by selecting the leading $n \times n$ sub-array, i.e. by a projection that deletes the last row and column. We discuss in section 3 what modifications are required to make the definition appropriate for contingency tables and log-linear models, where the issue of aggregation of levels also arises.

2.2 *Composition*

Let $\mathbf{n} = \{1, \ldots, n\}$, so that $\mathbf{n}^2 = \mathbf{n} \times \mathbf{n}$ is the set of ordered pairs of real numbers $\{(i, j): 1 \leq i, j \leq n\}$. It may be helpful to think of $\mathbf{n}^2$ as a table, or array. For the moment, this is an array of empty slots, but the intention is to fill it with real numbers as in a square matrix or contingency table. A function $f: \mathbf{n}^2 \to \mathcal{R}$ is a square matrix with components $f_{ij} \equiv f(i, j)$, each of which is a real number. Such a function is typically displayed as a square matrix, such as

$$f = \begin{pmatrix} 3.1 & 4.6 & 1.7 & 0.5 \\ 3.2 & 5.6 & 3.3 & 1.5 \\ 3.7 & 1.8 & 2.8 & 1.2 \\ 2.9 & 4.1 & 5.7 & 2.3 \end{pmatrix} \tag{2}$$

for $n = 4$.

Consider now the injective map $\varphi: \mathbf{m} \to \mathbf{n}$, in which $m = 3$ and $n = 4$, defined by

$$\varphi(1) = 4, \quad \varphi(2) = 2, \quad \varphi(3) = 1.$$

To say that the map is injective is to say that it is one-to-one, i.e. distinct points in the domain $\mathbf{m} = \{1, 2, 3\}$ have distinct images in $\mathbf{n} = \{1, 2, 3, 4\}$. With f defined as in (2), the composition $\varphi^* f$ is a square matrix of order 3 whose components are the values of f on the re-arranged sub-array

$$\varphi : \mathbf{m} \times \mathbf{m} \mapsto \begin{pmatrix} (4,4) & (4,2) & (4,1) \\ (2,4) & (2,2) & (2,1) \\ (1,4) & (1,2) & (1,1) \end{pmatrix}.$$

That is to say, $\varphi^* f$ is the 3×3 matrix

$$\varphi^* f = \begin{pmatrix} 2.3 & 4.1 & 2.9 \\ 1.5 & 5.6 & 3.2 \\ 0.5 & 4.6 & 3.1 \end{pmatrix}.$$

Note that diagonal elements remain on the diagonal, and off-diagonal elements remain off-diagonal.

It is clear that

$$\varphi^*(\alpha f + \alpha' f') = \alpha \varphi^* f + \alpha' \varphi^* f'$$

for all scalars α, α', showing that the induced transformation $\varphi^*\colon \mathcal{R}^{n^2} \to \mathcal{R}^{m^2}$ is a linear transformation from matrices of order n into matrices of order m. Note that the direction of φ^* is reversed relative to φ.

2.3 *Representation*

Denote by $\mathcal{I}$ the set of all injective maps on finite sets. It is sufficient for our purposes to pretend that, for each integer $n \geq 0$, there is only one set of size n, namely $\mathbf{n} = \{1, \ldots, n\}$. To each map $\varphi\colon \mathbf{n} \to \mathbf{n}'$ there coresponds a domain $\mathbf{n} = \operatorname{dom} \varphi$ and a codomain $\mathbf{n}' = \operatorname{cod} \varphi$: the image $\varphi \mathbf{n} \subset \mathbf{n}'$ is the set of points $\varphi x \in \mathbf{n}'$ such that $x \in \mathbf{n}$. It is important in all that follows not to confuse the image with the codomain. Two maps $\varphi\colon \mathbf{n} \to \mathbf{n}'$ and $\psi\colon \mathbf{n}' \to \mathbf{n}''$ are said to be composable if $\operatorname{dom} \psi = \operatorname{cod} \varphi$, in which case the composition $\psi\varphi\colon \mathbf{n} \to \mathbf{n}''$ is also a map. Each identity map $\mathbf{n} \to \mathbf{n}$ is evidently injective, and the composition of composable injective maps is also injective. The set $\mathcal{I}$ of injective maps on finite sets constitutes a category, containing each identity and closed under composition.

A contravariant representation of $\mathcal{I}$ by surjective linear maps is a function, here denoted by $\varphi \mapsto \varphi^*$, that associates with each 1–1 map $\varphi\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$ a surjective linear map $\varphi^*\colon \mathcal{V}_m \to \mathcal{V}_n$ on vector spaces in such a way that identity and composition are preserved. That is to say, the identity map $1\colon \mathbf{n} \to \mathbf{n}$ is carried to the identity linear transformation $1^*\colon \mathcal{V}_n \to \mathcal{V}_n$. Also, $\psi\colon \mathbf{n} \to \mathbf{n}'$ is carried to $\psi^*\colon \mathcal{V}_{n'} \to \mathcal{V}_n$, and the composition $\psi\varphi\colon \mathbf{m} \to \mathbf{n}'$ to

$$(\psi\varphi)^* = \varphi^* \psi^*\colon \mathcal{V}_{n'} \to \mathcal{V}_m,$$

which is the composition of images in reverse order.

The standard zero-order representation of $\mathcal{I}$ associates with each $\varphi\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$, the identity map $1\colon \mathcal{R} \to \mathcal{R}$. The *first-order standard representation of* $\mathcal{I}$ associates with each $\varphi\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$, the surjective linear transformation $\varphi^*\colon \mathcal{R}^n \to \mathcal{R}^m$ by functional composition $\varphi^* f = f \circ \varphi$. The components of the transformed vector are $(\varphi^* f)_i = f_{\varphi(i)}$, which is a selected subset, or sample, of the original components. Our interest is in square matrices, and specifically in vector spaces of square matrices. Accordingly, the *second-order standard representation of* $\mathcal{I}$ associates with each $\varphi\colon \mathbf{m} \to \mathbf{n}$ in $\mathcal{I}$, the surjective linear transformation $\varphi^*\colon \mathcal{R}^{n^2} \to \mathcal{R}^{m^2}$, also by functional composition

$$(\varphi^* f)_{ij} = f_{\varphi(i)\varphi(j)}$$

on square matrices as described and illustrated in the preceding section.

The action of an injective map on a table or multi-way array of numbers is to select, permute and re-label factor levels, for example by selecting varieties in a variety trial or

treatment levels in a horticultural experiment. Statistical sampling is an instance of the action of an injective map, in fact a randomly chosen injective map whose image is the set of sampled units. Ordinarily, in factorial designs the injective maps act independently on the levels of each factor. For a statistical design with k logically unrelated factors, a linear model is sub-representation of the product category $\mathcal{I}^k$ in the kth-order tensor product of first-order representations. The structure of these sub-representations is precisely that of the factorial models, also called hierarchical interaction models, in 1–1 correspondence with the free distributive lattice on k generators. But, for homologous factors (McCullagh, 2000), the action on the rows is the same as the action on the columns, and the matrices are square. That is to say, the structure of the sub-representations in the second-order standard representation of $\mathcal{I}$ is different from the sub-representations of the product category $\mathcal{I}^2$ in the tensor product of first-order representations. For example, symmetry and skew-symmetry and quasi-symmetry are sub-representations, but these are not factorial models.

2.4 *Orbits*

The symmetric group S_n acts on $\mathbf{n}$ by permutation. The action is said to be transitive because, to each ordered pair of points (i, j) there corresponds a permutation $\pi \in S_n$ such that $\pi(i) = j$. The symmetric group S_n also acts on the square array $\mathbf{n} \times \mathbf{n}$ by permutation, the same permutation applied to rows as to columns. In other words, the ordered pair (i, j) is carried to $\big(\pi(i), \pi(j)\big)$. This action has two orbits, the diagonal elements (i, i), and the off-diagonal elements (i, j) for $i \neq j$. In the category of injective maps, each injective map $\varphi\colon \mathbf{m} \to \mathbf{n}$ also preserves group orbits. The diagonal orbit $\mathrm{diag}(\mathbf{m}^2)$ is carried injectively to $\mathrm{diag}(\mathbf{n}^2)$, and similarly for the off-diagonal orbit. The orbit partition implies that the second-order standard representation of $\mathcal{I}$ may be decomposed as the direct sum of two sub-representations, the real-valued functions on the diagonal orbit and the real-valued functions on the off-diagonal orbit.

Ordinarily, there is no guarantee that group orbits are preserved by the non-invertible maps in the category. For example in the category of all maps on finite ordinal sets, the invertible maps are the finite symmetric groups. So the group orbits are the diagonal and the off-diagonal. But this partition is not preserved by non-injective maps: an ordered pair (i, j) with $i \neq j$ may be carried to an ordered pair for which $\varphi(i) = \varphi(j)$. On the other hand, diagonal elements cannot be sent to off-diagonal elements, so the diagonal is a sink, not an orbit.

2.5 *Sub-representation*

Let $\varphi\colon \mathbf{m} \to \mathbf{n}$ be a generic injective map, and let $\varphi^*\colon \mathcal{R}^{n^2} \to \mathcal{R}^{m^2}$ be the image map in the standard representation. Within this representation there may exist a sequence of subspaces $\mathcal{V}_n \subset \mathcal{R}^{n^2}$ such that, for each $\varphi\colon \mathbf{m} \to \mathbf{n}$, the linear transformation $\varphi^*\colon \mathcal{R}^{n^2} \to \mathcal{R}^{m^2}$ also satisfies $\varphi^*\mathcal{V}_n = \mathcal{V}_m$. Then the restriction of φ^* to the subspaces $\mathcal{V}_n$ also constitutes a representation of $\mathcal{I}$ by surjective linear maps on vector spaces. By a slight abuse of terminology, we say that $\mathcal{V}$ is a sub-representation in the standard representation, the maps being determined by restriction.

Several sub-representations have previously been indicated. The partition of $\mathbf{n}^2$ by orbits implies that the second-order standard representation contains two complementary sub-representations, the real-valued functions taking the value zero on the diagonal orbit, and

the real-valued functions taking the value zero on the off-diagonal orbit. We now introduce a model-formula notation capable of describing these and some others in a moderately convenient manner.

$$
\begin{aligned}
1 &= \{f \mid f_{ij} = a\} & 1^* &= \{f \mid f_{ij} = a;\ f_{ii} = 0\} \\
A &= \{f \mid f_{ij} = a_i\} & A^* &= \{f \mid f_{ij} = a_i;\ f_{ii} = 0\} \\
B &= \{f \mid f_{ij} = a_j\} & B^* &= \{f \mid f_{ij} = a_j;\ f_{ii} = 0\} \\
\mathrm{sym}(A,B) &= \{f \mid f_{ij} = f_{ji}\} & \mathrm{sym}^*(A,B) &= \{f \mid f_{ij} = f_{ji};\ f_{ii} = 0\} \\
\mathrm{diag}(A,B) &= \{f \mid i \neq j \Rightarrow f_{ij} = 0\} & \mathrm{alt}(A,B) &= \{f \mid f_{ij} = -f_{ji}\}
\end{aligned}
$$

The definition for A is to be read in the following manner. To each integer n there corresponds a vector subspace $A_n \subset \mathcal{R}^{n^2}$ consisting of $n \times n$ matrices f such that $f_{ij} = a_i$ for some vector $a \in \mathcal{R}^n$. The asterisk denotes restriction to the off-diagonal orbit; diag() denotes restriction to the diagonal orbit. For each $n \geq 0$, the vector spaces A_n and B_n have dimension n. For each $n \geq 2$, the vector spaces A_n^* and B_n^* also have dimension n. The representations $A \cong B$ are isomorphic, as are $A^* \cong B^*$. The representations A and A^* are almost, but not quite, isomorphic.

2.6 *Group representations*

The set of invertible maps $\mathbf{n} \to \mathbf{n}$ in $\mathcal{I}$ is the permutation group, and the representation theory for finite groups is well understood; see, for example, James and Liebeck (1993) or Diaconis (1988). Since the restriction of each $\mathcal{I}$-representation to $\mathbf{n}$ is necessarily a group representation, we begin our study of $\mathcal{I}$-representations by studying the group representations that occur in the standard representation. The great majority of group representations are not inherited under restriction, and thus do not extend naturally to $\mathcal{I}$-representations. Our task is to identify those that do extend naturally.

A representation of S_n associates with each permutation π a linear transformation $\pi^*\colon \mathcal{V} \to \mathcal{V}$ in such a way that identity and composition are preserved. In the first-order standard representation, $\mathcal{V} = \mathcal{R}^n$ and π^* is the usual $n \times n$ matrix representation by linear transformations that permute coordinates. This representation decomposes into two irreducible representations, the trivial one-dimensional representation $1_n \subset \mathcal{R}^n$ by constant functions, and a complementary representation $1_n^\perp$ by vectors whose components add to zero. The restriction of 1_n to a subset of $m < n$ components is equal to 1_m, so the sequence $1_n \subset \mathcal{R}^n$ is a sub-representation of $\mathcal{I}$. However, the restriction of $1_n^\perp$ is equal to $\mathcal{R}^m$, not to $1_m^\perp$. Thus, the first-order standard $\mathcal{I}$-representation contains exactly one sub-representation, but there is no complementary sub-representation. In this respect, the structure of sub-representations of $\mathcal{I}$ is fundamentally different from the representation theory for finite groups: it is not semi-simple.

In the second-order standard representation $\mathcal{V} = \mathcal{R}^{n^2}$, and the linear transformation π^* acts by composition $(\pi^* f)_{ij} = f_{\pi(i)\pi(j)}$. The group product $\pi\varphi$ is carried to $(\pi\varphi)^* = \varphi^*\pi^*$, reversing the order of composition.

Character theory for finite groups (James and Liebeck 1993) enables us to identify irreducible sub-representations that occur in a given representation. In the case of the symmetric group S_n, to each partition of the number n there corresponds a distinct irreducible representation, and conversely, so each irreducible sub-representation is associated with a partition such as $(n), (n-1,1), (n-2,2), (n-1,1,1)$ and so on. The decomposition

of the first-order representation is $(n) \oplus (n-1,1)$ in which (n) is the trivial representation and $(n-1,1)$ is the $(n-1)$-dimensional representation. For the standard second-order representation of S_n by square matrices, the decomposition by sub-representations is as follows.

$$\begin{aligned}
\mathcal{R}^{n^2} &\cong \mathrm{diag}(\mathcal{R}^{n^2}) \oplus \mathrm{sym}^*(\mathcal{R}^{n^2}) \oplus \mathrm{alt}(\mathcal{R}^{n^2}) \\
\mathrm{diag}(\mathcal{R}^{n^2}) &\cong \mathcal{R}^n \cong (n) \oplus (n-1,1) \\
\mathrm{sym}^*(\mathcal{R}^{n^2}) &\cong (n) \oplus (n-1,1) \oplus (n-2,2) \\
\mathrm{alt}(\mathcal{R}^{n^2}) &\cong (n-1,1) \oplus (n-2,1,1)
\end{aligned}$$

The first line states that the vector space of square matrices can be expressed as the direct sum of three subspaces, the diagonal matrices, the symmetric off-diagonal matrices, and the skew-symmetric matrices. Further, these three subspaces are preserved under coordinate permutation, which is to say that each is a sub-representation of S_n. None of these sub-representations is irreducible. The diagonal sub-representation contains a one-dimensional sub-representation by constant functions plus a complementary $(n-1)$-dimensional sub-representation consisting of functions whose components sum to zero. The symmetric off-diagonal representation contains a one-dimensional representation by constant functions, a sub-representation consisting of functions

$$f_{ij} = \alpha_i + \alpha_j \qquad (i \neq j)$$

such that $\sum \alpha_i = 0$, and a complementary sub-representation

$$(n-2,2) \cong \{f \mid f_{ij} = f_{ji}, \quad f_{ii} = 0, \quad \sum_i f_{ij} = 0\}.$$

The dimensions are 1, $n-1$ and $n(n-3)/2$. The alternating representation contains two sub-representations, one consisting of functions

$$f_{ij} = \alpha_i - \alpha_j$$

and a complementary sub-representation

$$(n-2,1,1) \cong \{f \mid f_{ij} = -f_{ji}, \quad \sum_i f_{ij} = 0\}.$$

The dimensions are $n-1$ and $(n-1)(n-2)/2$. All dimensions are necessarily non-negative, so the expression $n(n-3)/2$ is interpreted as zero for $n \leq 3$.

It is straightforward to verify that each of these subspaces is closed under coordinate permutation. The marvel of character theory is that it enables us to determine whether or not a given representation is reducible. These calculations are standard and classical, so details are omitted. For $n \geq 3$, the standard second-order S_n-representation contains seven irreducibles,

$$\mathcal{R}^{n^2} \cong (n)^{\oplus 2} \oplus (n-1,1)^{\oplus 3} \oplus (n-2,2) \oplus (n-2,1,1)$$

in which the trivial representation (n) has multiplicity two, $(n-1,1)$ has multiplicity three, and the remaining two have multiplicity one each.

2.7 *Inheritance and $\mathcal{I}$-representations*

A sub-representation of $\mathcal{I}$ in the standard representation is a sequence of subspaces $\mathcal{V}_n \subset \mathcal{R}^{n^2}$ such that

(i) $\mathcal{V}_n$ is a representation of the symmetric group S_n, closed under coordinate permutation;

(ii) The restriction of functions f in $\mathcal{V}_n$ to the leading $m \times m$ sub-array is equal to $\mathcal{V}_m$.

Thus, each component of an $\mathcal{I}$-representation is also a group representation, but only certain group representations can occur as a component in an $\mathcal{I}$-representation.

A representation $\mathcal{V}$ that contains no sub-representation other than itself and the zero representation is called *irreducible*. For the category $\mathcal{I}$ of injective maps, or for the product category $\mathcal{I}^k$, such representations are rather rare, so this concept is not especially useful. A representation that is expressible as the direct sum $\mathcal{V} \cong \mathcal{U} \oplus \mathcal{W}$ of two complementary non-zero sub-representations is called *decomposable*. Conversely, a representation that is not expressible in this form is called *indecomposable*. The standard first-order $\mathcal{I}$-representation is indecomposable, but it is not irreducible because it contains the trivial one-dimensional sub-representation. A representation that is expressible as the span $\mathcal{V} = \mathcal{U} + \mathcal{W}$ of two sub-representations, not equal to zero or $\mathcal{V}$, is called *weakly decomposable*. Although $\mathcal{V} \cap \mathcal{W}$ is also a sub-representation, there need not exist a direct-sum decomposition. The factorial model $A + B$ is a weakly decomposable $\mathcal{I}^2$-representation in which $A \cap B$ is the trivial one-dimensional sub-representation. There is no direct-sum decomposition, so this representation is $\mathcal{I}^2$-indecomposable.

It is evident that the direct-sum decomposition

$$\mathcal{R}^{n^2} \cong \mathrm{diag}(\mathcal{R}^{n^2}) \oplus \mathrm{sym}^*(\mathcal{R}^{n^2}) \oplus \mathrm{alt}(\mathcal{R}^{n^2})$$

is preserved not only by permutation but also by restriction to subsets. This observation implies that the standard $\mathcal{I}$-representation decomposes as the direct sum of three sub-representations. This decomposition is unique only in the sense of isomorphism. There exist alternative direct-sum-decompositions in which $\mathrm{diag}(\mathcal{R}^{n^2})$ is replaced by an isomorphic sub-representation such as A, B or $\mathrm{symadd}(A, B)$. Despite this lack of uniqueness, we focus initially on the sub-representations that occur in each of these three representations. On account of isomorphisms, however, there exist sub-representations of the standard second order representation that are not in any of these three components. Many of these are indecomposable, and some are irreducible.

The method used to construct a sub-representation proceeds as follows. For some large n, choose a group irreducible $\mathcal{V}^{(n)} \subset \mathcal{R}^{n^2}$. Then act on $\mathcal{V}^{(n)}$ by restriction to the leading $m \times m$ sub-array giving a vector space $\mathcal{V}_m^{(n)}$, which is necessarily a sub-representation of the group S_m in $\mathcal{R}^{m^2}$. The sequence of vector spaces $\mathcal{V}_m^{(n)}$ is a sub-representation of the category $\mathcal{I}^{(n)}$ of injective maps on finite sets of cardinality $\leq n$. To extend this finite sequence to an infinite sequence, it is necessary to begin at $n = \infty$, whatever that might mean, and to define the sequence $\mathcal{V}_m^{\infty}$ by restriction to finite squares. The representation generated in this manner may contain a sub-representation, i.e. it may not be irreducible, but it is necessarily indecomposable, not expressible as the vector span of two non-zero sub-representations.

To see how this works, let $\mathcal{V}^{(n)} = (n) \subset \mathrm{diag}(\mathcal{R}^{n^2})$ be the trivial group representation by constant diagonal matrices, multiples of the identity. The restriction of $\mathcal{V}^{(n)}$ to the leading $m \times m$ sub-array is equal to $\mathcal{V}^{(m)} \subset \mathrm{diag}(\mathcal{R}^{m^2})$, which is also S_m-irreducible. Thus, the

constant diagonal functions 1_d constitute a trivial $\mathcal{I}$-representation. On the other hand, if we let $\mathcal{V}^{(n)} \cong (n-1,1) \subset \mathrm{diag}(\mathcal{R}^{n^2})$ be the complementary group irreducible consisting of diagonal matrices whose elements sum to zero, the restriction to a proper $m \times m$ sub-array is equal to the set of all $m \times m$ diagonal matrices. The zero-sum condition is not preserved by restriction to subsets. In other words, the restriction of the S_n-irreducible $(n-1,1)$ is equal to $(m) \oplus (m-1,1)$, which is clearly not irreducible. That is to say, the sequence of vector subspaces $\mathrm{diag}_n \subset \mathcal{R}^{n^2}$ is the smallest $\mathcal{I}$-representation that includes the group irreducibles $(n-1,1) \subset \mathrm{diag}(\mathcal{R}^{n^2})$. This representation contains a one-dimensional trivial sub-representation, but there does not exist a complementary sub-representation.

In the same way, we find that sym^* and alt are indecomposable $\mathcal{I}$-representations containing sub-representations as follows:

$$1^* \subset \mathrm{symadd}^* \subset \mathrm{sym}^*; \qquad \mathrm{altadd} \subset \mathrm{alt}\,.$$

For $n \geq 2$, 1^*_n is the one-dimensional vector space of constant off-diagonal matrices. Likewise, symadd^*_n is the vector subspace consisting of matrices expressible in the form

$$f_{ij} = \alpha_i + \alpha_j \quad (i \neq j)$$

for some $\alpha \in \mathcal{R}^n$. Equivalently, symadd^*_n is the image of the linear transformation $\mathcal{R}^n \to \mathcal{R}^{n^2}$ as defined above. The dimension is zero for $n \leq 1$, one for $n = 2$, and n for $n \geq 3$. Finally, altadd_n is a vector space of dimension $n-1$ consisting of additive skew-symmetric matrices of the form $\alpha_i - \alpha_j$. The mappings $\mathcal{R}^n \to \mathcal{R}^{n^2}$ defined by $(g_n\alpha)_{ij} = \alpha_i + \alpha_j$ for $i \neq j$ and $(h_n\alpha)_{ij} = \alpha_i - \alpha_j$ are the components of two homomorphisms of the first-order representation into the second-order standard representation.

2.8 *General sub-representations*

When applied to a group irreducible, $\mathcal{V}^{(n)}$, the projection $\mathcal{V}^{(n)}_m = \varphi^*\mathcal{V}^{(n)}_n$ is a representation of S_m in $\mathcal{R}^{m^2}$. The sequence $\{\mathcal{V}^{(n)}_m\}$ is a representation of the category $\mathcal{I}^{(n)}$ by surjective linear maps. When applies to the direct sum of S_n-irreducibles, the projection yields

$$\varphi^*(\mathcal{V}^{(n)} \oplus \mathcal{W}^{(n)}) = \mathcal{V}^{(n)}_m + \mathcal{W}^{(n)}_m$$

in which the image spaces may have non-zero intersection. Since we eventually let n go to infinity, it is sufficient to consider only the sequences generated by restriction of S_n-irreducibles. Such representations are necessarily indecomposable in both senses. All other sub-representations of $\mathcal{I}$ in the standard representation are expressible as the span of indecomposables.

Let θ, ϕ be arbitrary real numbers. Consider the S_n-irreducible consisting of vectors f such that, for some $\alpha \in 1^\perp_n$

$$f_{ij} = \begin{cases} \phi\alpha_i & \text{if } i = j \\ \alpha_i \cos\theta + \alpha_j \sin\theta & \text{otherwise.} \end{cases}$$

For $n \geq 3$, and for each θ, ϕ, this irreducible has dimension $n-1$ and is isomorphic with $1^\perp_n \cong (n-1,1)$. If $\phi = 0$ and $\theta = -\pi/4$, the restriction to $m < n$ elements is an irreducible

S_m-representation consisting of skew-symmetric matrices of the form $\alpha_i - \alpha_j$, and this sequence evidently constitutes an irreducible $\mathcal{I}$-representation. For every other value of θ, ϕ, the restriction is a representation of S_m of the same form except that the zero-sum restriction on α is not conserved. These representations are almost isomorphic with the first-order standard representation, and are indecomposable but not irreducible.

The unusual qualifier 'almost' is inserted here for the following reason. Let $\phi = 0$ and $\theta = \pi/4$, so that $f_{ij} = \alpha_i + \alpha_j$ for $i \neq j$. The dimension of this representation is zero for $n = 1$, one for $n = 2$ and n for each $n \geq 3$, whereas the dimension of the first-order standard representation is n. On dimensional grounds alone, the representations cannot be isomorphic in the orthodox sense. However, if we restrict the category $\mathcal{I}$ to sets of size at least three, the truncated representations are indeed isomorphic. The phrase 'almost isomorphic' is understood in this truncated sense.

To each real ϕ and $-\pi/2 < \theta \leq \pi/2$ there corresponds an indecomposable $\mathcal{I}$-representation $\mathcal{V}_{\theta,\phi}$. For $(\theta, \phi) \neq (\theta', \phi')$ the representations $\mathcal{V}_{\theta,\phi}$ and $\mathcal{V}_{\theta',\phi'}$ may overlap. For example, if $\phi = 1$, the representations corresponding to $\theta = 0$ and $\theta = \pi/2$ are the subspaces $f_{ij} = \alpha_i$ and $f_{ij} = \alpha_j$, conventionally denoted by A and B. The intersection is the one-dimensional trivial representation. However, the representation $A + B$ is in fact $\mathcal{I}$-decomposable in the strong sense because it can be expressed as symadd $\oplus$ altadd. The span of any subset of the $\mathcal{V}_{\theta\phi}$ is evidently a decomposable representation, but any subset of three distinct $\mathcal{V}$'s is sufficient for a basis.

This completes the story for the three isomorphic S_n-representations. If we do the same thing for the representation $(n-1, 1, 1)$, with $n \geq 3$ we find that the projection for $m < n$ is the space of all skew-symmetric matrices. This representation is $\mathcal{I}$-indecomposable, but it contains the additive sub-representation of dimension $n - 1$. Likewise, the projection of the symmetric representation $(n - 2, 2)$ is the space of symmetric off-diagonal matrices, which contains two sub-representations. Finally, the two trivial sub-representations 1 and 1^* are non-overlapping and almost isomorphic.

2.9 *Symmetry*

It is possible to reduce the number of representations by adding further symmetry conditions such as invariance under the two-element group S_2 that acts by switching rows with columns. While this invariance may seem appealing on purely algebraic grounds, I have rarely seen compelling arguments for it in statistical applications. Nonetheless, it is worthwhile explaining why certain $\mathcal{I}$-representations in the standard representation do not extend to representations of $\mathcal{I} \times S_2$.

The inversion in the group S_2 carries (i, j) to (j, i), so the orbit decomposition is unaffected. Since f_{ij} is carried to f_{ji}, the decomposition by diagonal, symmetric and skew-symmetric matrices is unaffected. The group S_2 acts as the identity on the symmetric matrices, $(f^T = f)$, but the inversion acts as the negative identity on skew-symmetric matrices, $(f^T = -f)$. Consequently, the additive symmetric sub-representation, $f_{ij} = \alpha_i + \alpha_j$ is carried to itself element-wise. In the additive skew-symmetric representation, however, the inversion carries $f_{ij} = \alpha_i - \alpha_j$ to $-f$, so these representations are no longer isomorphic. In effect, S_2 splits the three previously isomorphic representations into two isomorphic representations and one other. The inversion in S_2 carries the subspace $A_\theta = \{f = \alpha_i \cos\theta + \alpha_j \sin\theta\}$ to $A_{\pi/2-\theta}$, so A_θ is a representation only for $\theta = -\pi/4$ and $\theta = \pi/4$. In particular, the row factor $A \equiv A_0$ and the column factor $B \equiv A_{\pi/2}$ are not

representations of $\mathcal{I} \times S_2$. However $A+B \cong A_{\pi/4} \oplus A_{-\pi/4}$ is a decomposable representation.

Quasi-symmetry and quasi-skew-symmetry are both decomposable sub-representations of $\mathcal{I} \times S_2$:

$$QS \cong \mathrm{diag} \oplus \mathrm{sym}^* \oplus \mathrm{altadd},$$
$$QSS \cong \mathrm{alt} \oplus \mathrm{symadd}^* .$$

3. Contingency tables

3.1 *Aggregation of levels*

The discussion in section 2 relates to representation-theory for injective maps, the idea being that the form of the model, i.e. the model formula, should be unaffected by permutation and selection of factor levels. This condition seems fairly compelling in certain circumstances, particularly in connection with explanatory factors having unordered levels. For contingency tables, however, certain factors are explanatory whereas others are responses. It is extremely important to make this distinction clear because the relevant algebraic operations on response levels are not the same as the operations performed on the levels of an explanatory factor. The levels of a response factor may be aggregated, but the same operation does not make sense for explanatory factors unless the levels being aggregated are equivalent in their effect. Selection of response levels may also be appropriate when we talk of conditional distributions. For these reasons, it is appropriate in connection with contingency tables to develop the representation theory for all maps, including aggregation and selection, along the lines of section 2.

To illustrate what is involved, consider a response factor initially having levels denoted by $\Omega = \{a, b, c\}$, in which levels $\{a, b\}$ are aggregated to form a new binary response factor having two levels. This operation is denoted by a map $\varphi\colon \Omega \to \Omega'$ such that

$$\varphi(a) = \varphi(b) = 1, \quad \varphi(c) = 2,$$

and $\Omega' = \{1, 2\}$. Let $f = (f(a), f(b), f(c))$ be a vector of frequencies, or a measure defined on the subsets of Ω. Then $f \circ \varphi^{-1}$ is the measure on Ω' whose density is

$$f \circ \varphi^{-1}(1, 2) = \big(f(a) + f(b), f(c)\big).$$

This is a linear transformation $\varphi^\dagger\colon \mathcal{R}^3 \to \mathcal{R}^2$ for which the matrix is

$$\varphi \mapsto \varphi^\dagger = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that the direction of $\varphi^\dagger$ is the same as the direction of φ, so this association determines a covariant representation. We focus instead on the dual, or contravariant, representation in which the matrix is replaced by its transpose and φ^* is the dual of $\varphi^\dagger$.

3.2 *Representation theory*

Consider the category $\mathcal{A}$ of all maps on finite sets, i.e. injective maps, surjective maps and compositions of these. The standard representation A of order one associates with each set Ω the vector space $A_\Omega = \mathcal{R}^\Omega$, and with each map $\varphi\colon \Omega \to \Omega'$ the linear map $\varphi^*\colon \mathcal{R}^{\Omega'} \to \mathcal{R}^\Omega$ by functional composition. That is to say, for each $f \in \mathcal{R}^{\Omega'}$, the image $\varphi^* f$ is given by composition $(\varphi^* f)(i) = f(\varphi(i))$. Note that if φ is injective, φ^* is surjective, and if φ is surjective φ^* is injective. The standard representation contains the one-dimensional trivial representation $1 \subset A$ of constant functions. The restriction of the standard $\mathcal{A}$-representation to injective maps coincides with the $\mathcal{I}$-representation described in section 2.

For the standard representation of order two, however, the structure of the sub-representations of $\mathcal{A}$ is not exactly the same as the structure of the sub-representations of $\mathcal{I}$ in A^2. In particular, there is no partition into two orbits, and there is no sub-representation corresponding to diagonal matrices. This is fairly simple to understand. Let $f = \mathrm{diag}\{1, 2, 3, 4\}$ be the diagonal matrix of order four, and let $\varphi\colon \mathbf{3} \to \mathbf{4}$ be the map

$$\varphi(1) = \varphi(2) = 3; \quad \varphi(3) = 1.$$

Then $\varphi^* f$ is the $\mathbf{3} \times \mathbf{3}$ matrix

$$\varphi^* f = \begin{pmatrix} 3 & 3 & 0 \\ 3 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

which is not a diagonal matrix. On the other hand, if f is zero on the diagonal, $\varphi^* f$ is also zero on the diagonal, so the off-diagonal matrices constitute a sub-representation in A^2, which splits into symmetric and skew-symmetric off-diagonal matrices. There is in fact a complementary representation isomorphic with A, but this is not unique. In the direct-sum decomposition

$$A^2 \cong A \oplus \mathrm{sym}^* \oplus \mathrm{alt},$$

A may be identified with the row factor, the column factor, or with the space of additive symmetric matrices. Regardless of the choice, $1 \subset A$ is a one-dimensional trivial sub-representation consisting of matrices whose elements are all equal. Furthermore, alt contains an additive sub-representation, but by contrast with $\mathcal{I}$-representations, sym^* is $\mathcal{A}$-irreducible.

If $\mathcal{U}, \mathcal{V}$ are sub-representations of $\mathcal{A}$ in a given representation $\mathcal{X}$, the intersection and vector span are also sub-representations. Consequently, the entire set of sub-representations in A^2 constitutes a lattice, as shown in Figure 1 below. The penultimate row lists the three irreducible sub-representations. For $\theta \neq -\pi/4$, the sub-representation $A_\theta \cong A$ consists of matrices expressible in the form $f_{ij} = a_i \cos\theta + a_j \sin\theta$ for some vector $a \in A$. Then $A \equiv A_0$ and $B \equiv A_{\pi/2}$ are the row factor and the column factor respectively. In any event, symmetry, quasi-symmetry, and quasi-skew-symmetry are all sub-representations.

For tables whose rows and columns are indexed by unrelated factors with unordered levels, the factorial models are the only models that are closed under permutation and selection of levels. What we aim to do here is to provide a similar list of the models that are suitable for square contingency tables with rows and columns indexed by the same factor. The models displayed in Fig. 1 comprise precisely such a list.

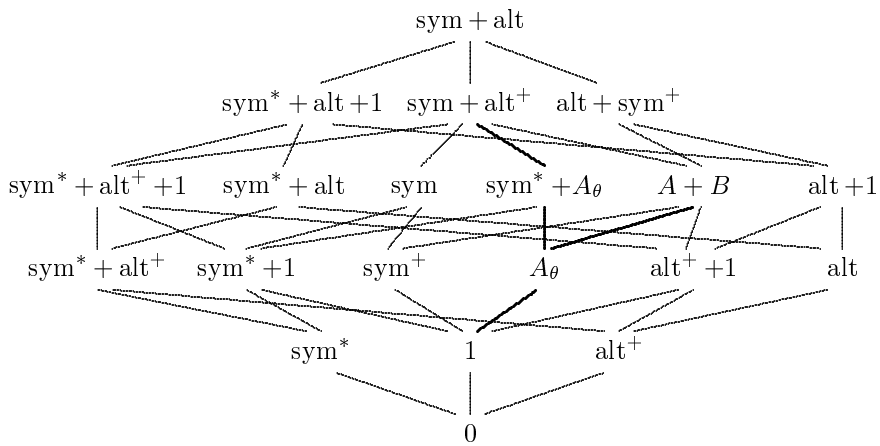

Figure 1: *Hasse diagram of the sub-representations of $\mathcal{A}$ in A^2. The off-diagonal symmetric sub-representation is* sym^**;* sym *is the full symmetric subspace;* alt *is the full skew-symmetric subspace;* sym^+ *and* alt^+ *are the additive symmetric and alternating subspaces; the remaining subrepresentations isomorphic with A have been telescoped into one, denoted by A_θ. Quasi-symmetry is the sub-representation* $\mathrm{sym} + \mathrm{alt}^+$

3.3 *Aggregation and log-linear models*

Let $\mathrm{vct}(\Omega^2)$ be the vector space of square matrices, and let $\mathrm{vct}^+(\Omega^2)$ be the cone of non-negative square matrices. It is helpful for present purposes to regard $\mathrm{vct}^+(\Omega^2)$ as the set of all non-negative measures on Ω^2, and this set includes the observed contingency table. To each map $\varphi\colon \Omega \to \Omega'$ in $\mathcal{A}$ there corresponds a map $\varphi^\dagger\colon \mathrm{vct}(\Omega^2) \to \mathrm{vct}(\Omega'^2)$ as described in section 3.1 by composition with the inverse image. Suppose, for example, that $\Omega = \{1,2,3,4\}$, $\Omega' = \{1,2,3\}$ and

$$\varphi(1) = \varphi(2) = 1; \quad \varphi(3) = 2; \quad \varphi(4) = 3.$$

Then the 4×4 contingency table in $\mathrm{vct}(\Omega^2)$

$$\begin{pmatrix} 7 & 4 & 3 & 1 \\ 2 & 9 & 5 & 3 \\ 6 & 2 & 6 & 2 \\ 0 & 7 & 3 & 8 \end{pmatrix} \xmapsto{\varphi^\dagger} \begin{pmatrix} 22 & 8 & 4 \\ 8 & 6 & 2 \\ 7 & 3 & 8 \end{pmatrix}$$

is carried by aggregation to the 3×3 table in $\mathrm{vct}(\Omega'^2)$ as shown above. Note that the direction of $\varphi^\dagger$ is the same as the direction of φ. Note that $\mathrm{vct}^+(\Omega^2) \subset \mathrm{vct}(\Omega^2)$ is carried to $\mathrm{vct}^+(\Omega'^2) \subset \mathrm{vct}(\Omega'^2)$ by the linear transformation $\varphi^\dagger$, which is in fact the dual, or matrix transpose, of φ^*. By the same token, the subset of symmetric product distributions such that $F(A \times B) = G(A)G(B)$ is preserved by this aggregation operation. These are both sub-functors in $\mathrm{vct}(\Omega^2)$, and their intersection is also a sub-functor. They are not sub-representations because $\mathrm{vct}^+(\Omega^2)$ is not a vector space, nor is the set of product distributions a vector space.

Consider a contravariant representation Θ, one of the points in the lattice in Fig. 1. For each set Ω in $\mathcal{A}$, exponential weighting by $\theta \in \Theta_\Omega$ carries the distribution $F \in \mathrm{vct}^+(\Omega^2)$ to a new distribution F_θ such that $dF_\theta(x) = e^{\theta(x)}\, dF(x)$ for each $x \in \Omega^2$, or equivalently,

$$\log\left(\frac{dF_\theta}{dF}\right) = \exp(\theta).$$

Since we are not considering probability distributions here, re-normalization is not necessary.

Exponential weighting is an instance of a natural transformation g of functors as illustrated below.

$$\begin{array}{ccccc} \Omega & & \mathrm{vct}^+(\Omega^2) \times \Theta_\Omega & \xrightarrow{g_\Omega} & \mathrm{vct}^+(\Omega^2) \\ \varphi\big\downarrow & & \varphi^\dagger\big\downarrow \quad \big\uparrow\varphi^* & & \big\downarrow\varphi^\dagger \\ \Omega' & & \mathrm{vct}^+(\Omega'^2) \times \Theta_{\Omega'} & \xrightarrow{g_{\Omega'}} & \mathrm{vct}^+(\Omega'^2) \end{array} \tag{3}$$

By this diagram we mean that the component g_Ω of the natural transformation g is exponential weighting of non-negative distributions $g_\Omega(F, \theta) = F_\theta$. To say that these are the components of a natural transformation is to say that for each $F \in \mathrm{vct}(\Omega^2)$ and $\theta \in \Theta_{\Omega'}$, the induced transformations are such that

$$g_{\Omega'}(\varphi^\dagger F,\, \theta) = \varphi^\dagger g_\Omega(F,\, \varphi^*\theta).$$

Note that $\varphi^\dagger F$ is the marginal distribution of F after aggregation or transformation by φ. The commutativity condition is a consistency condition to the effect that, if the parameter θ is constant over the levels to be aggregated, marginalization and exponential weighting can be performed in either order. This consistency property requires Θ to be a representation of $\mathcal{A}$; it is not sufficient that Θ be a representation of $\mathcal{I}$. See the remark concerning the avoidance of inconsistencies in section 7 of Stigler's paper in the present volume.

A log-linear model for a square contingency table is determined by the set of mean measures, which is a subset of $\mathrm{vct}^+(\Omega^2)$. The parameter space for a log-linear model has two components, a set of baseline measures $\Phi_\Omega \subset \mathrm{vct}^+(\Omega^2)$ and a representation $\Theta_\Omega \subset \mathcal{R}^{\Omega^2}$. The set of baseline measures is a sub-functor, and exponential weighting gives rise to an enlarged set of measures as follows.

$$\begin{array}{ccccc} \Omega & & \Phi_\Omega \times \Theta_\Omega & \xrightarrow{g_\Omega} & \mathrm{vct}^+(\Omega^2) \\ \varphi\big\downarrow & & \varphi^\dagger\big\downarrow \quad \big\uparrow\varphi^* & & \big\downarrow\varphi^\dagger \\ \Omega' & & \Phi_{\Omega'} \times \Theta_{\Omega'} & \xrightarrow{g_{\Omega'}} & \mathrm{vct}^+(\Omega'^2) \end{array} \tag{4}$$

That is to say, the log-linear model is the set of measures in the image of g_Ω. Note that the baseline sets Φ_Ω are measures closed under aggregation, but the image of g is not closed under aggregation. For example, a symmetric matrix remains symmetric under aggregation, but a quasi-symmetric matrix does not remain quasi-symmetric.

It is natural to ask why a set of baseline measures is required. Why not use the uniform distribution on Ω or Ω^2 as a baseline. The answer is that the uniform distribution does not remain uniform under aggregation of levels, so aggregation and exponential weighting do not commute.

Generally speaking, the model as described above is over-parameterized, which is to say that g_Ω is many-to-one. Thus, only certain functions of the parameter are identifiable. Furthermore, the same model may be identified in more than one manner. Thus, for example, quasi-symmetry requires $\Theta = QS$, but the baseline set can be either the set of symmetric measures, the set of symmetric product measures, or the set of product measures. These baseline measures are sub-functors in vct^+ and are thus non-negative.

4. Conclusions

Because of the need to incorporate a set of baseline measures in (4), it may well be the case that two distinct representations Θ, Θ' give rise to the same image, i.e. the same model with a different parameterization. Let Φ be the set of symmetric product measures, Φ' the set of product measures, Φ'' the set of symmetric measures, and $\Theta = QS$. Regardless of the choice of Φ, the log density with respect to uniform measure on Ω^2 has the form

$$\log \pi_{ij} = \alpha_i + \beta_j + \phi_{ij}$$

where ϕ_{ij} is symmetric. The set of log-linear models that can be generated from Φ as baseline is the set of representations in Fig. 1 that includes sym^+, as follows

$$\mathrm{sym}^+, \quad \mathrm{sym}, \quad A+B, \quad QS, \quad QSS, \quad \mathrm{sym}+\mathrm{alt} \equiv A.B$$

the final model being saturated. The conclusion is that, for a square contingency table indexed by homologous factors, there are exactly six log-linear models with Φ as baseline. The fact that one of these is quasi-symmetry can hardly be a surprise. But why quasi-skew-symmetry?

Since $\Phi \subset \Phi'$ and $\Phi \subset \Phi''$, the set of models that can be generated from Φ' or Φ'' is a subset of the preceding list. Other than marginal homogeneity (MH) and measures concentrated on the diagonal, it does not appear that there is any other sub-functor in vct^+, closed under aggregation, that could serve as a baseline. When marginal homogeneity and diagonal measures are are included in the list, the number of models increases to eight. Among these eight models, six are in fact sub-functors in vct^+, which is to say closed under aggregation of levels. The log-linear models QS and QSS satisfy the commutativity condition (4), but they are not closed under aggregation of levels, so they are not functors in vct^+. For example, the 4×4 table

$$\begin{pmatrix} 50.24 & 68.71 & 27.20 & 55.85 \\ 20.71 & 28.33 & 19.29 & 27.67 \\ 444.20 & 353.29 & 240.52 & 323.99 \\ 193.85 & 220.67 & 159.99 & 215.50 \end{pmatrix}$$

is quasi-skew-symmetric in the log-linear sense because $m_{ij}m_{ji} = m_{ii}m_{jj}$ up to rounding error. But the 3×3 table formed by aggregating the second and third rows and columns is not quasi-skew-symmetric. In this sense quasi-skew-symmetry is similar to quasi-symmetry.

The situation for multi-way tables indexed by k unrelated factors is similar so far as the commutative diagram (4) is concerned. In this situation, however, we must allow for permutation, selection, or aggregation, independently on the levels of each factor. The natural minimal baseline model Φ in vct^+ is the set of product measures, i.e. joint distributions for which all factors are independent. The set of log-linear models generated from this baseline is then in 1–1 correspondence with the factorial models, also called hierarchical interaction models, that include all k main effects. For a two-way table, the only log-linear models are $A+B$ and $A.B$. For a three-way table we have $A+B+C$, $A.B+C[3]$, $A.B + B.C[3]$, $A.B + B.C + A.C$ and $A.B.C$, making a total of nine models. With the exception of the three conditional independence models and the one indecomposable model, the other five are closed under aggregation of levels. The special status that is sometimes accorded to decomposable log-linear models or graphical models (Lauritzen, 1996) is not a consequence of representation theory.

References

Agresti, A. (1983) Testing marginal homogeneity for ordinal categorical variables. *Biometrics* **39**, 505–510.

Barry, D. and Hartigan, J.A. (1993) Choice models for predicting divisional winners in major league baseball. *J. Am. Statist. Assoc.* **88**, 766–774.

Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W. (1975) *Discrete Multivariate Analysis.* MIT Press.

Bradley, R.A. and Terry, M.E. (1952) Rank analysis of incomplete block designs I: The method of paired comparisons. *Biometrika* **39**, 324–345.

Caussinus, H. (1965) Contribution à l'analyse statistique des tableaux de correlation. *Ann. Fac. Sci. Univ. Toulouse* **29**, 77–182.

Diaconis, P. (1988) *Group Representations in Probability and Statistics.* IMS Lecture Notes **11**, Hayward, CA.

Ewens, W.J. and Spielman, R.S. (1995) The transmission/disequilibrium test: history, subdivision and admixture.

Fienberg, S.E. and Larntz, K. (1976) Loglinear representation for paired and multiple comparison models. *Biometrika* **63**, 245–254.

James, G. and Liebeck, M. (1993) *Representations and Characters of Groups.* Cambridge University Press.

Joe, H. (1990) Extended use of paired comparisons models, with application to chess rankings. *Appl. Statist.* **39**, 85–93.

Kelly, F. (1979) *Reversibility and Stochastic Networks.* Chichester, Wiley.

Lauritzen, S.L. (1996) *Graphical Models.* Clarendon Press, Oxford.

McCullagh, P. (1982) Some applications of quasi-symmetry. *Biometrika* **69**, 303–308.

McCullagh, P. (2000) Invariance and factorial models (with discussion). *J. R. Statist. Soc.* B **62**, 209–256.

McCullagh, P. (2001) What is a statistical model? *Ann. Statist.* (to appear).

Pike, M.C., Casagrande, J. and Smith, P.G. (1975) Statistical analysis of individually matched case-control stidies in epidemiology: Factor under study a discrete variable taking multiple values. *Br. J. Prev. Soc. Med.* **29**, 196–201.

Stewart, J.Q. (1948) Demographic gravitation: Evidence and applications. *Sociometry* **1**, 31–58.

Sethuraman, B. and Speed, T. (1997) Transmission disequilibrium test via marginal homogeneity. (Unpublished manuscript).

Spielman, R.S., McGinnis, R.E. and Ewens, W.J. (1993) Transmission test for linkage disequilibrium. *Amer. J. Human Genetics* **52**, 506–16.

Stigler, S. (1994) Citation patterns in the journals of statistics and probability. *Statistical Science* **9**, 94–108.

Stigler, S. (2001) The missing early years of contingency tables.

Upton, G.J.G. (1985) Modelling cross-tabulated regional data. In *Measuring the Unmeasurable* Ed. P. Nijkamp. NATO Scientific Publications, 1985, 197–218.

Wermuth, N. (1987) Parametetric collapsibility and the lack of moderating effects in contingency tables with a dichotomous response variable. *J. R. Statist. Soc.* B **49**, 353–364.

Whittemore, A. (1978) Collapsibility of multi-dimensional contingency tables. *Biometrics* **27**, 545–562.