Factorization Models and Other Representations of Independence

by
Robert J. Buehler and
Yong Goo Lee[1]
University of Minnesota
Technical Report No. 496

August 1987

Key words: Hierarchical model, log linear models, factorization models, graphical models, decomposable models, recursive models, ZPA models, conditional independence.

Direct correspondence to: Professor Robert J. Buehler, School of Statistics, 270 Vincent Hall, Minneapolis, MN 55455.

---

[1] Partially supported by NIH Grant GM 25271.

Abstract

Factorization models are a generalization of hierarchical log linear models which apply equally to discrete and continuous distributions. In regular (strictly positive) cases the conjunction of two factorization models is another factorization model whose representation is obtained by a simple algorithm. Failure of this result in an irregular case is related to a theorem of Basu on ancillary statistics. It is shown how factorization models are related to ZPA (zero partial association), graphical, decomposable and recursive models.

Key words. Hierarchical model, log linear models, factorization models, graphical models, decomposable models, recursive models, ZPA models, conditional independence

1. Introduction

The purpose of this paper is review methods of representing independence in multivariate distributions and to relate them to factorization models, a generalization of hierarchical log linear models. Many of the key ideas of factorization models are implicit, for example, in Wermuth (1976a), (1980), Dawid (1979a), (1980b), Darroch et al (1980), and Kiiveri et al (1984), Lauritzen et al (1984).

In Section 2 we define factorization models, and establish that the intersection of two factorization models is again a factorization model. This furnishes the basis of a "factorization calculus" for routine manipulations. The failure of the calculus for irregular cases is shown to be related to a problem with a theorem of Basu on ancillary statistics.

In Section 3 factorization conditions are compared with other representations of independence, including "Dawid conditions," ZPA (zero partial association), graphical, decomposable and recursive models. Although factorization and Dawid conditions partition distributions differently, the partitions become identical for saturated graphical factorizations and saturated Dawid conditions ("saturated" means involving all variables). It is further argued that not all recursive models are factorization models, but every decomposable factorization model is a recursive model.

The way in which factorization models generalize hierarchical log linear models can be seen by an example. A p.m.f. $f(x,y,z)$ is expressible in the factored form $a(x,y)\ b(x,z)$ iff the hierarchical log linear model lacks terms u_{23} and u_{123} (in the notation of Bishop et al (1975)). Equivalent notations are: The "fitted marginals" are {AB}, {AC} (Goodman (1970)), the "generating class" is {{1,2}, {1,3}} (Haberman (1974), Darroch et al (1980)), the "sufficient configuration" is C_{12}, C_{13} (Bishop et al (1975)). Fienberg (1977) abbreviated further to [12][13] and Wermuth (1976 a,b) to 12/13.

2. Factorization Models

In what follows the distributions can be either discrete or continuous. Our treatment is non-measure theoretic and assumes conditional densities to be defined as the quotient of joint and marginal densities.

For a trivariate p.m.f. or p.d.f. $f(x,y,z)$ Dawid (1979a) pointed out that

(2.1) $\quad X \perp\!\!\!\perp Y | Z$ iff $f(x,y,z) = a(x,z)b(y,z)$.

Example 2.1. If $f(x_1, \ldots, x_6) = ce^{-Q/2}$, $Q = \Sigma\Sigma a_{ij}x_ix_j$, a_{13}

$= a_{14} = a_{23} = a_{24} = 0$, then $(X_1, X_2) \perp\!\!\!\perp (X_3,X_4) | (X_5,X_6)$. This result can be obtained by noting that the conditional covariance matrix is the inverse of a block diagonal four-by-four submatrix of $\{a_{ij}\}$, but a vector generalization of (2.1) establishes the result by factorization, avoiding matrices. (For general results on multivariate normal independence structures, see Wermuth (1976a) and Speed and Kiiveri (1986).)

2.1 Definitions

Def. 2.1. The order of a model, N, is the number of joint random variables.

Def. 2.2. A factor is a number 1,...,N. The set $\{1,\ldots,N\}$ will be denoted by $\mathcal{N}$.

This terminology agrees with Darroch et al (1980).

Def. 2.3. Any subset of $\mathcal{N}$, say $a \subseteq \mathcal{N}$, will be called a term.

Def. 2.4. A product $\mathcal{A}$ is a set of terms (not necessarily distinct).

We will variously write for example

$$\mathcal{A} = \{a,b,c\} = \{[1],[12],[23]\} = 1/12/23, \tag{2.2}$$

which is a mix of Haberman, Fienberg and Wermuth notation. After reduction as defined below our product corresponds to Haberman's "generating class." The separate term "product" is retained so as to include the nonminimal case, and to suggest the factorization of f into a product.

Def. 2.5. Reduction of a product means deletion of all terms which are proper subsets of other terms and deletion of all but one of any duplicated terms. For example, reduction of {[1],[12],[23],[12]} yields {[12],[23]}.

Def. 2.6. A product is minimal if it has no reduction.

Def. 2.7. Two products are equivalent, $\mathcal{A} \approx \mathcal{B}$ if they reduce to the same minimal product.

Def. 2.8. The class $C_{\mathcal{A}}$ is the set of functions f which factor in accordance with $\mathcal{A}$.

It is sometimes helpful to think of $f \in C_{\mathcal{A}}$ as equivalent to log f belonging to linear subspace. In notation close to that of Darroch, Lauritzen and Speed (1980) p. 524, and Darroch and Speed (1983) p. 725, $f \in C_{\mathcal{A}}$ iff

(2.3) $$\log f = \Sigma_{a \in \mathcal{A}} \xi_a(x_a)$$

where x_a is the set of x_i for $i \in a$. In this case Darroch and Speed (1983) write $\log f \in M_{\mathcal{A}}$.

Example 2.2. If $\mathcal{A} = \{[12],[234],[345]\}$, then $f \in C_{\mathcal{A}}$ iff there exist a,b,c such that

$$f(x_1,\ldots, x_5) = a(x_1,x_2)b(x_2,x_3,x_4)c(x_3,x_4x_5).$$

Proposition 2.1. For the class of strictly positive p.m.f.s or p.d.f.s., the sets $C_{\mathcal{A}}$ for all minimal $\mathcal{A}$ correspond one-to-one with hierarchical log linear models whose generating classes (in the sense of Haberman (1974)) are $\mathcal{A}$. Consequently the number of distinct factorizations equals the number of hierarchical models.

Set operations on terms pose no special difficulties but special conventions are useful for products.

Def. 2.9. Set operations on products.

$\mathcal{A} \subseteq \mathcal{B}$ means every term in $\mathcal{A}$ is in $\mathcal{B}$.

$\underline{\mathcal{A}} \leq \mathcal{B}$ means every $a \in \mathcal{A}$ is a subset of some $b \in \mathcal{B}$.

$\mathcal{A} \wedge \mathcal{B}$ means the set of mn terms $c_{ij} = a_i \cap b_j$, $i = 1,\ldots m$, $j = 1,\ldots,n$, where $\mathcal{A} = \{a_1 \ldots a_n\}$, $\mathcal{B} = \{b_1 \ldots b_n\}$

The notation $\mathcal{A} \wedge \mathcal{B}$ agrees with Lauritzen et al (1984), p.16.

It is trivial to show:

Proposition 2.2. $\mathcal{A} \subseteq \mathcal{B}$ implies $\mathcal{A} \leq \mathcal{B}$, and $\mathcal{A} \leq \mathcal{B}$ implies $C_{\mathcal{A}} \subseteq C_{\mathcal{B}}$.

2.2. Factorization Calculus.

A more general version of (2.1) is:

Proposition 2.3. (The Dawid-factorization connection.) If a,b,c are a partition of $\mathcal{N} = \{1,\ldots,N\}$ and X,Y,Z are corresponding vector variates then

$$(2.4)\quad X \perp\!\!\!\perp Y|Z \quad \text{iff} \quad f \in C_{\mathcal{A}}, \quad \mathcal{A} = \{a \cup c,\ b \cup c\}.$$

The case c = null set can be accomodated by agreeing that $X \perp\!\!\!\perp Y|Z$ then becomes $X \perp\!\!\!\perp Y$.

Example 2.3. Let f(0,0.0) = f(1,1,1) = 1/3, f(1,0,0) = f(0,1,1) = 1/6, f=0 otherwise. Then $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ but $X \perp\!\!\!\perp (Y,Z)$ is false.

When do $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ imply $X \perp\!\!\!\perp (Y,Z)$? For $\mathcal{A} = \{[13],[23]\}$ write $C_{\mathcal{A}} = C_{13/23}$, etc. Then in factorization notation the question translates to: Does $C_{13/23} \cap C_{12/23} =$

$C_{1/23}$? Dawid (1980b), Sec. 6 and 7, gives a measure theoretic treatment.

Proposition 2.4. For N = 3 if f is strictly positive, then $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ iff $X \perp\!\!\!\perp (Y,Z)$.

In the discrete case a proof can be given by determining which terms are zero in the log linear expansion of log f. Alternatively we can write

$$(2.5) \qquad f = (f_{13}/f_3)f_{23} = (f_{12}/f_2)f_{23},$$

exhibiting two marginal-conditional factorizations. Under the regularity assumptions, $f_{23} \neq 0$ and so $f_{13}/f_3 = f_{12}/f_2$. The LHS is free of y and the RHS is free of z. Thus both sides depend on x only and the result follows easily. In the Appendix we show how this approach extends to certain irregular cases and how it relates to Basu's (1982) Theorem 2.

The hierarchical log linear approach can be extended to continuous models via difference operators. Let us define

$$\begin{aligned}
&\Delta_1 f(x,y,z) = f(x',y,z) - f(x,y,z),\\
(2.6) \qquad &\Delta_{12} f(x,y,z) = f(x',y',z) - f(x,y,z),\\
&\Delta_1\Delta_2 f(x,y,z) = f(x',y',z) - f(x',y,z) - f(x,y',z) + f(x,y,z),
\end{aligned}$$

etc. It is known that for $N = 3$ there are 19 hierarchical log linear models. By including permutations, these are obtainable from the list of generating classes $\mathcal{A}$ below. It is straightforward to verify that $f \in C_{\mathcal{A}}$ iff the corresponding difference operator operating on f gives zero. For the discrete case the averaging operators of Darroch and Speed (1983) p. 729, provide an alternative characterization.

Generating class	Difference operator
ϕ	Δ_{123}
1	Δ_{23}
1/2	$\Delta_3, \Delta_1\Delta_2$
1/2/3	$\Delta_{12}\Delta_{13}\Delta_{23}$
12	Δ_3
1/23	$\Delta_1\Delta_{23}$
12/13	$\Delta_2\Delta_3$
12/13/23	$\Delta_1\Delta_2\Delta_3$
123	ϕ

In general we obtain difference operators from a generating class $\mathcal{A} = \{a_1 \ldots a_k\}$ as follows: Put $a = \cup_j a_j$, $c = \mathcal{N} \backslash a$, $b_j = a \backslash a_j$. Then

(2.6) $f \in C_A$ iff $\Delta_c f = 0$ and $\Delta_{b_1} \dots \Delta_{b_k} f = 0$.

For routine manipulations the following generalization of Proposition 2.4 is useful.

Proposition 2.5. (Factorization calculus.) For any class of strictly positive functions f, $C_A \cap C_B = C_{A \wedge B}$

Proof. By definition, $A \wedge B \leq A$ and $A \wedge B \leq B$. By Proposition 2.2, $C_{A \wedge B} \subseteq C_A$ and $C_{A \wedge B} \subseteq C_B$, and so $C_{A \wedge B} \subseteq C_A \cap C_B$. The converse is less obvious, and does require some regularity in view of Example 2.3. Needed are lemmas like: if $a_1(x,y)a_2(z,w) = a_3(x,z)\, a_4(y,w)$ then both products equal $b_1(x)b_2(y)b_3(z)b_4(w)$. This is hardly surprising, and a proof can be given by applying the difference operators mentioned above to log f. The conditions $f \in C_A$ and $f \in C_B$ give two sets of difference equations equal to zero. Standard algebraic techniques give the required combined set difference equations.

2.3 Examples.

Example 2.4 By Proposition 2.3, $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ translate to $A = 13/23$ and $B = 12/23$, giving $A \wedge B = 1/3/2/23 \approx 1/23$, which translates to $X \perp\!\!\!\perp (Y,Z)$, showing Proposition 2.4 to be a special case of 2.5.

Example 2.5. (Markov chain.) Assuming $X_1 \perp\!\!\!\perp (X_3,X_4)|X_2$ and $(X_1,X_2) \perp\!\!\!\perp X_4|X_3$ gives $\mathcal{A} = 12/234$, $\mathcal{B} = 123/34$, $\mathcal{A} \wedge \mathcal{B} = 12/23/34$. One explicit factorization is the marginal-conditional:

$$(2.7) \qquad f_{1234} = f_{12}(f_{23}/f_2)(f_{34}/f_3) = f_{12}f_{3.2}f_{4.3}.$$

Example 2.6. (Exponential family.) Let $f(x,y,\alpha,\beta) = C(\alpha,\beta)h(x,y)e^{\alpha x+\beta y}$. This represents an exponential family with α,β fixed parameters. For our purposes imagine a joint prior density of (α,β) incorporated in the term $C(\alpha,\beta)$. With numbering 1,2,3,4 for α,β,x,y the factorization 12/34/13/24 is evident. By Proposition 2.5 this is equivalent to $C_{124/134} \cap C_{123/234}$, which translates by Proposition 2.3 to $X \perp\!\!\!\perp \beta|(Y,\alpha)$ and $Y \perp\!\!\!\perp \alpha|(X,\beta)$. In the terminology of Dawid (1975), Basu (1977) and Barndorff-Nielsen (1978), X is specific sufficient for α, Y is specific sufficient for β, x is specific ancillary for β, and y is specific ancillary for α. One explicit factorization of an arbitrary $f \in C_{12/34/13/24}$ is incidentally

$$(2.8) \qquad \frac{f_{1234}}{f_{....}} = \frac{f_{12..}}{f_{1...}}\,\frac{f_{1.3.}}{f_{..3.}}\,\frac{f_{.2.4}}{f_{.2..}}\,\frac{f_{..34}}{f_{...4}}\,,$$

where $f_{12..}(x_1,x_2,x_3,x_4) = f_{1234}(x_1,x_2,c_3,c_4)$, etc. In contrast to Example 2.5 it is impossible here to factor into marginal and conditional p.m.f.s. (Andersen (1974), Wermuth (1976a) p. 102, Wermuth (1980) p. 967, Darroch et al. (1980), p. 528; Proposition 3.3 below).

Example 2.7 (ZPA conditions). Take N = 5. Wermuth (1976a) writes ZPA (1,2) if $1 \perp\!\!\!\perp 2 \mid (3,4,5)$ for which the factorization representation is 1345/2345. The ZPA manipulations of Wermuth (1976a) p 254-5, are a special case of the present factorization calculus. To see this, apply Proposition 2.5 to Wermuth's example of finding the conjunction of ZPA(1,2), ZPA(1,3) and ZPA (2,3). We find: (1345/2345) $\wedge$ (1245/2345) = 145/2345, and (145/2345) $\wedge$ (1345/1245) = 145/245/345.

2.4. The Factorization Partition

Given a family F of distributions a set of m conditions (equivalenty m models--for example, if A is model then $f \in C_A$ is a condition), the conditions potentially partition F into 2^m sets, but some may be empty. In the case of factorization conditions there is a drastic reduction due to the hierarchical structure.

Proposition 2.6. For any family F of strictly positive distributions in N dimensions, factorization conditions partition F into at most H(N) nonempty sets, where H(N) is the number of hierarchical log linear models of order N.

Proof. Given $f \in F$, define $P(f) = \{\mathcal{A} | f \in C\}$, and $\mathcal{A}_{min} = \mathcal{A}_{min}(f) = \bigwedge_{\mathcal{A} \in P} \mathcal{A}$ (with obvious reference to the notation $\mathcal{A} \wedge \mathcal{B}$). By Proposition 2.5, $f \in C_{\mathcal{A}}$, and $\mathcal{A}_{min} \leq \mathcal{A}$ for all $\mathcal{A} \in P(f)$. Let $\mathcal{A} < \mathcal{B}$ denote $\mathcal{A} \leq \mathcal{B}$ and $\mathcal{A} \neq \mathcal{B}$, and put $D_{\mathcal{A}} = C_{\mathcal{A}} - \bigcup_{\mathcal{B}<\mathcal{A}} C_{\mathcal{B}}$. We will show that $\{D_{\mathcal{A}}\}$ is a partition of F. (i) Since $f \in D_{\mathcal{A}_{min}}$ we have $\cup D_{\mathcal{A}} = F$. (ii) Assume $f \in D_{\mathcal{A}} \cap D_{\mathcal{B}}$, for some $\mathcal{A} \neq \mathcal{B}$. It follows that $f \in C_{\mathcal{A}} \cap C_{\mathcal{B}} = C_{\mathcal{C}}$ where $\mathcal{C} = \mathcal{A} \wedge \mathcal{B}$. Since $\mathcal{A} \neq \mathcal{B}$ we have either $\mathcal{A} < \mathcal{C}$ or $\mathcal{B} < \mathcal{C}$. It follows that either $f \notin D_{\mathcal{A}}$ or $f \notin D_{\mathcal{B}}$, which contradicts $f \in D_{\mathcal{A}} \cap D_{\mathcal{B}}$.

The number of partition sets $D_{\mathcal{A}}$ equals H(N), the number of hierarchical log linear models.

Darroch et al. (1980) page 537 state that H(1) = 2, H(2) = 5, H(3) = 19, H(4) = 167, H(5) = 7580.

Proposition 2.6 is false without the strictly positive assumption. The f given in Example 2.2 belongs to $C_{12/23}$ and $C_{13/23}$ but not to $C_{1/23}$, and it belongs to both $D_{12/23}$ and $D_{13/23}$. Accordingly $\{D_{\mathcal{A}}\}$ is not always a partition when F includes arbitrary f's. The partition induced by taking all unions and intersections of $\{C_{\mathcal{A}}\}$ in these unrestricted cases will generally have more than H(N) nonempty partition sets.

2.5 Saturated Factorization Models

As is well known, hierarchical models involve uniformity as well as independence conditions. For example, with $N = 2, A = \{[1]\}$, $f \in C_A$ means $X \perp\!\!\!\perp Y$ and Y has a uniform distribution. Hierarchical or factorization models are brought closer to the models of Section 3 below by restricting to "saturated" models:

Def.2.10. A generating class $A = \{a_1, \cdots a_k\}$ is saturated if $\cup a_j = N$.

The family of saturated factorization models has the same closure property (Proposition 2.5) as the unrestricted family. Similarly Proposition 2.6 continues to hold: The number of partition sets equals the number of models (for example, 9 for N = 3).

3. Relationships to Other Models

Factorization models provide one method of classification. In this section we take a brief look at the relationship to other classification schemes:

3.1 Dawid Conditions

In view of the efforts of Dawid (1979 a, b, 1980 a, b) to popularize his notation $X \perp\!\!\!\perp Y|Z$ we will refer to it as Dawid notation even though it was anticipated by Goodman (1970) and no doubt others as well.

The definition of what constitutes a Dawid condition requires an arbitrary choice. We choose to allow $X_a \perp\!\!\!\perp X_b|X_c$ where a, b, c are disjoint subsets of $\mathcal{N} = \{1,\ldots N\}$ ("terms"). If either a or b is null, the condition is empty; if c is null we understand (as previously stated) the condition to be $X_a \perp\!\!\!\perp X_B$.

Definition 3.1. If $a \cup b \cup c = \mathcal{N}$ the condition $X_a \perp\!\!\!\perp X_b \mid X_c$ is called saturated. Otherwise it is unsaturated.

By Proposition 2.3, every saturated Dawid condition is a factorization condition. For N=3 the unsaturated $X \perp\!\!\!\perp Y$ is not a factorization condition, and the factorization 12/13/23 is known not to be representable by a combination of Dawid conditions. Thus for N > 2 the factorization and Dawid conditions yield partitions not ordered by inclusion. The relationship is studied further in Proposition 3.2 and the remarks which follow.

The find the Dawid partition for N = 3, first list all conditions as $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp (Y,Z)$ and their cyclic permutations. Since $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Z|Y$ iff $X \perp\!\!\!\perp (Y,Z)$, the latter and its permutations are not needed. A minimal set of six generators is $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$ and permutations. These potentially yield $2^6 = 64$ partition sets. That not all are occupied is implied by additional relationships such as Proposition 2.4.

<u>Proposition 3.1</u>. The Dawid partition has cardinality 18 for the family of full support I by J by K contingency tables.

<u>Proof</u>. The partition sets can be identified by a binary code in which 0 = not satisfied, 1 = satisfied, and the six conditions are ordered $X \perp\!\!\!\perp Y$, $Y \perp\!\!\!\perp Z$,...$Z \perp\!\!\!\perp X|Y$. Then 000100 for example denotes $X \perp\!\!\!\perp Y|Z$ satisfied and the other five conditions fail. The 18 nonull partition sets are ("3" in parentheses denotes permutation multiplicity): 000000, 100000(3), 000100(3), 110000(3), 100100(3), 110110(3), 111000, 111111. A more detailed proof is given by Lee (1986) and Lee and Buehler (1986).

For I by J by 2 tables it is known (Birch, 1963) that $X \perp\!\!\!\perp Y$

and $X \perp\!\!\!\perp Y|Z$ imply either $X \perp\!\!\!\perp Z$ or $Y \perp\!\!\!\perp Z$, which means 100100 is impossible, reducing the count to 17 (010010 and 001001 remain possible because of the preferred Z direction).

For N = 4 it can be shown that permutations of the following are generators: $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp Y|(Z,W)$. We do not know the cardinality.

3.2 Graphical Models.

As mentioned above, factorization models are one-to-one with hierarchical models. A subclass of hierarchical models are the decomposable ("Markov") models of Goodman (1970, 1971) and Haberman (1974). Intermediate between hierarchical and decomposable models are the graphical models of Darroch et al (1980).

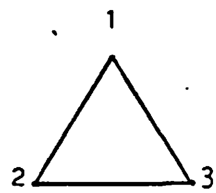
Let us associate with each factor a vertex. Given any generating class A, construct a graph as follows: two vertices are joined with an (undirected) line (an "edge") iff the corresponding factors occur together in any term. Such a pair of vertices are called adjacent or neighbors. A set of vertices is a complete subset if all pairs of the set are neighbors. A clique is a maximal complete subset.

Certain conventions are needed for unsaturated models. For $N = 4$, $A = \{[1], [23]\}$, vertices 2 and 3 are joined by an edge, vertex 1 stands alone, and vertex 4 is absent from the graph.

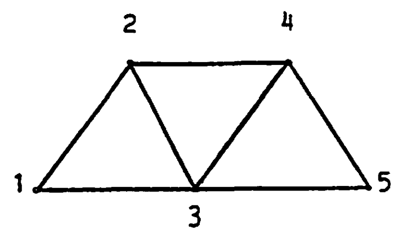
By the above construction any generating class determines a graph, and that graph determines (and is determined by) its set of cliques. Can the cliques be used to recover the generating class? Sometimes, but not always.

Def. 3.2. A generating class is graphical if the cliques it defines are the same as its terms.

Example 3.1. Consider generating classes $A = 123$, $B = 12/13/23$, $C = 123/234/345$. Both A and B have the graph



and a single clique [123], corresponding to the single term in A, but different from B. Thus A is graphical but B is not. The graph of C is



from which it is seen that the cliques are [123], [234], [345], the same as the terms so that C is graphical.

Among models of order 5, C defines one of 1450 graphical models and one of 7580 hierarchical models (Darroch et al. (1980), Table 3). Darroch et al display the graphs which, with their permutations, describe all graphical models of order 5.

As is known, graphical models give a convenient means of reading out independence conditions. Let a,b,c be disjoint terms such that $a \cup b \cup c = \mathcal{N}$. Suppose that no vertex in a is linked by an edge to a vertex in b. Then deletion of the c vertices separates the a set from the b set. This happens if $f \in C_{\mathcal{A}}$, $\mathcal{A} = \{a \cup c, b \cup c\}$ which by Proposition 2.3 is equivalent to $X_a \perp\!\!\!\perp X_b | X_c$. The argument works both ways.

Since a saturated graphical model has N vertices it has $\binom{N}{2} = N(N-1)/2$ possible edges. Since each edge is either present or absent, there are $2^{N(N-1)/2}$ saturated graphical models of order N. As indicated by Darroch et al (1980) the count of saturated plus unsaturated graphical models is

$$\sum_{i=0}^{N} \binom{N}{i} 2^{\binom{i}{2}} .$$

Saturated graphical models are generated by "zero partial association" or ZPA conditions (Wermuth (1976 a,b), Wermuth and Lauritzen (1983)). Following Wermuth and Lauritzen we will write (as in Example 2.7 above)

(3.1) ZPA (r,s) means $r \perp\!\!\!\perp s | \mathcal{N} \backslash r \backslash s$

For a model of order N there are $\binom{N}{2} = N(N-1)/2$ ZPA conditions. Each condition can be either satisfied or not satisfied, giving a partition of the space of functions f into $2^{N(N-1)/2}$ sets. How are these related to the $2^{N(N-1)/2}$ saturated graphical models of the previous paragraph? Extending slightly results of Wermuth (1976a) we have:

<u>Proposition 3.2.</u> Let $\{C_A\}$ be the set of saturated graphical factorization models (Defs. 2.11 and 3.2) and let $\{D_A\}$ be the corresponding partition sets as defined in the proof of Proposition 2.6, but restricted to the saturated graphical case. Then the D_A partition is identical with the ZPA partition.

<u>Proof</u>. Taken in conjunction with any graphical factorization model, the condition ZPA(r,s) deletes the (r,s) edge from the graph. Any ZPA model is characterized by a set I of pairs (r,s) in the triangular array $1 \le r < s \le N$. Define

$$Z^+(I) = \{f \mid ZPA(r,s) \text{ true for } (r,s) \in I\}$$
$$Z^-(I) = \{f \mid ZPA(r,s) \text{ false for } (r,s) \notin I\}$$
$$Z(I) = Z^+(I) \cap Z^-(I).$$

Let A be the generating class defined by the graph in which the (r,s) edge is present iff $(r,s) \notin I$, and let C_A and D_A be the corresponding factorization model and partition set. If $f \in Z(I)$, then $f \in Z^+(I)$, implying $f \in C_A$. But $f \in Z(I)$ also implies $f \in Z^-(I)$, which implies $f \notin C_B$ for any B strictly contained in A. From the definition of D_A it follows that $f \in Z(I)$

implies $f \in D_A$. Since both Z(I) and D_A define partitions, the result follows.

Example 3.2. Take N = 3. ZPA (1,2) true iff $1 \perp\!\!\!\perp 2 | 3$ iff $f \in C_{13/23}$. ZPA (1,2) true, ZPA (1,3) false and ZPA (2,3) false iff $f \in D_{13/23} = C_{13/23} \setminus C_{1/23} \setminus C_{2/13}$.

Example 3.3. N =3, ZPA (1,2) and ZPA (1,3) true iff $1 \perp\!\!\!\perp 2 | 3$ and $1 \perp\!\!\!\perp 3 | 2$ iff $1 \perp\!\!\!\perp 2,3$ iff $f \in C_{1/23}$. ZPA (1,2) and ZPA (1,3) true and ZPA (2,3) false iff $f \in D_{1/23} = C_{1/23} \setminus C_{1/2/3}$.

The point of Proposition 3.2 is this. For N > 2 the factorization and Dawid partitions differ for three reasons: (1) Nongraphical factorizations like 12/13/23 are nor representable by Dawid conditions. (2) Unsaturated factorizations like 1/2 also are not Dawid representable. (3) Unsaturated Dawid conditions like $X \perp\!\!\!\perp Y$ are not representable by factorization. But if we restrict to saturated graphical factorizations and to saturated Dawid conditions, then the resulting partitions are both the same as the ZPA partition.

3.3 Decomposable and Recursive Models.

The definition of decomposable (or Markov) generating class given by Haberman (1974) page 166 (or see Darroch et al (1980 page 524) is set theoretic and hence carries over to the present framework. Lauritzen et al (1984) have shown that every decomposable model is graphical, and that decomposability can be checked by inspecting the graph: the generating class is decomposable iff the graph contains no cycle of length ≥ 4 without a chord (is "triangulated"). (See also Darroch et al (1980)). The simplest cycle of length 4 appears above in Example 2.6: 12/24/43/31.

As is well known in the theory of contingency tables, decomposability is necessary and sufficient for existence of a closed form maximum likelihood estimate. To tie in with the recursive models let us adopt the notation of Kiiveri et al (1984) and write

$$(1234) = (12)(3.2)(4.13) \tag{3.2}$$

as an abbreviation for

$$f(x_1, \ldots, x_4) = f_{12}(x_1,x_2)f_{3.2}(x_3|x_2)f_{4.13}(x_4|x_1,x_3) \tag{3.3}$$

Either expression defines a recursive model which represents the family of distributions so expressible. As Kiiveri et al point out, either expression translates to the pair of Dawid conditions $3 \perp\!\!\!\perp 1|2$ and $4 \perp\!\!\!\perp 2|(1,3)$. Replacing (12) by the equivalent (1)(2.1), a general expression of this form would define a recursive model by

$$(3.4) \qquad (1)(2.d_1)\ (3.d_2)\ \ldots(N.d_{N-1})$$

where d_j is a (possible empty) subset of $\{1, \ldots, j\}$. If $e_j = \{1, \ldots, j\}\backslash d_j$, then the model is uniquely specified by giving either $d_1, \ldots, d_{N-1}$ or $e_1, \ldots, e_{N-1}$. For $3 \perp\!\!\!\perp 1|2$ the notation ZPD(3.1) is sometimes used (for example Wermuth and Lauritzen (1983)) denoting "zero partial dependence." For $r > s$, ZPD(r,s) means $r \perp\!\!\!\perp s|\{1, \ldots, r-1\}\backslash\{s\}$. The information in (3.4) can be replaced by a set of ZPD's within a triangular array of pairs: (2,1), (3,1), (3,2), (4,1), ..., (N,N-1). The number of ZPD conditions is $\binom{N}{2} = N(N-1)/2$ and the cardinality of the partition is $2^{N(N-1)/2}$.

Not all recursive models are factorization models, the simplest counterexample being (1)(2)(3.12). The model (3.2) is obviously contained in the factorization class $C_{12/23/134}$, but the converse is false (Goodman (1971), equations (4.6), (4.7)), so that (3.2) likewise is not a factorization model.

Wermuth and Lauritzen (1983) have defined reducible patterns of ZPD's and have shown that reducibility is necessary and sufficient for a recursive model to be graphical factorization model.

The expressions (3.2)-(3.4) may be called "factor at a time" recursive expressions. An alternative is vector ("term") at a time. Slightly adjusting the notation of Darroch et al (1980) p 529, we have

$$f(x) = f(x_{a_1}) \prod_{i=2}^{k} f(x_{b_i} | x_{c_i}) \tag{3.5}$$

$$= \frac{\prod_{i=1}^{k} f(x_{a_i})}{\prod_{i=2}^{k} f(x_{c_i})} \tag{3.6}$$

(There is a typo in Darroch et al (1980) where b replaces c in the last expresion.) For this to make sense we need $a_1, b_2, \ldots, b_k$ a partition of $N = \{1, \ldots, N\}$, $a_2 = b_2 \cup c_2, \ldots, a_k = b_k \cup c_k$, and $c_2 \subset a_1$, $c_3 \subset a_1 \cup a_2, \ldots, \quad c_k \subset a_1 \cup \ldots \cup a_{k-1}$. One point to note is that (3.5) can easily be put in the form (3.4) by breaking up individual terms, as in

$$(456.12) = (4.12)(5.124)(6.1245).$$

Moreover reversing this procedure poses no problems.

The main point we wish to make is the relationship of (3.5) (3.6) to factorization models. Using $b_i \subseteq a_i$, $i = 2, \ldots, k$, (3.6) gives trivially, $f \in C_A$, $A = \{a_1 \ldots a_k\}$. Does $f \in C_A$ conversely imply (3.6)? The argument of Darroch et al (1980), shows the role of the "decomposability" condition

$$(3.7) \qquad c_i \subset a_{r_i} \text{ for some } r_i \in \{1, \ldots, i-1\},\ i = 2, \ldots, k.$$

We have by inductively calculating terms in the product (3.5):

Proposition 3.3. If $f \in C_A$ where $= \{a_1 \ldots a_k\}$ satisfies (3.7), then f has the recursive form (3.5).

Here is is to be understood that $A = \{a_1 \ldots a_k\}$ is given initially, reordered if need be, and b_i, c_i are defined as in Darroch, et al (1980): The "new part" of each a_i is $b_i = a_i \cap (a_1 \cup \ldots \cup a_{i-1})$ and the "overlap" is $c_i = a_i \backslash b_i$.

From the above discussion we conclude that not all recursive models are factorization models, but every decomposable factorization model is a recursive model.

Appendix

Irregular Cases

Example 2.2 shows how the factorization calculus fails when there are zeros in the domain of f. In this appendix we will give necessary and sufficient conditions on the support of a discrete f for certain factorization results to hold. Similar results have been given by Basu (1958), Koehn and Thomas (1975), Bishop et al (1975) Chapter 5, and Dawid (1979b, 1980b). These papers are in part concerned with Basu's "Theorem 2" (see Basu (1982) for an overview of Theorems 1, 2, and 3 on sufficiency and ancillarity). Briefly the connection is as follows: Let T = sufficient statistic, U = ancillary statistic, θ = parameter, S = sufficiency condition expressed as $U \perp\!\!\!\perp \theta | T$, I = independence condition expressed as $U \perp\!\!\!\perp T | \theta$, A = ancillary condition expressed as $U \perp\!\!\!\perp \theta$. Basu's Theorem 2 states: S and I imply A, which follows from Proposition 2.4. Fuller discussions can be found in the references cited above.

Let f(x,y,z) be defined on a finite discrete set $S = S_x \times S_y \times S_z$. Let S_{yz} be the marginal support of y and z. Two points (y,z) and (y',z') in S_{yz} are called <u>y-linked</u> if y = y' and <u>z-linked</u> if z = z'. Two points are <u>chain linked</u> if they can be joined by a chain of y and z linked points.

Suppose there exist nontrivial partitions of S_y into $A \cup A^c$ and S_z into $B \cup B^c$ (where c denotes complement) such that S_{yz} is contained in $(AB) \cup (A^cB^c)$. Then the set $A \times B$ will be called an _yz splitting set_. (This terminology is adapted from Koehn and Thomas (1975). It is closely related to the concept of separability in Bishop et al. (1975) Section 5.4.2)

Proposition A.1. Every pair of points in S_{yz} is chained linked iff there does not exist a yz splitting set.

Starting with equation (2.5) we can show:

Proposition A.2. Assume $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$. Within any set of chain linked y,z points $f_{13}(x,z)/f_3(z)$ and $f_{12}(x,y)/f_2(y)$ depend on x only.

Proposition A.3. $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ imply $X \perp\!\!\!\perp (Y,Z)$ iff there does not exist a yz splitting set.

Proof. If there does not exist a splitting set then Proposition A.1 shows that all points are chain linked and Proposition A.2 shows that $f_{13}(x,z)/f_3(z)$ depends only on x, and can be called a(x). Thus $f(x,y,z) = a(x)f_{23}(y,z)$, showing $X \perp\!\!\!\perp (Y,Z)$. Example 2.3 shows $X \perp\!\!\!\perp (Y,Z)$ can fail when there is a splitting set.

REFERENCES

Andersen, A.H. (1974), "Multidimensional contengency tables," Skandinavian Journal of Statistics, 1, 115-121.

Barndorff-Nielsen, O. (1978), Information and Exponential Families in Statistical Theory, New York: John Wiley.

Basu, D. (1958), "On Statistics Independent of Sufficient Statistics," Sankhya, 20, 223-226.

Basu, D. (1982), "On the Elimination of Nuisance Parameters," Journal of the American Statistical Association, 72, 355-356.

Basu, D. (1982), "Basu Theorems," Encyclopedia of Statistical Sciences, Kotz, Johnson and Read, editors, Vol. 1, 193-196, New York: John Wiley.

Birch, N.W. (1963), "Maximum Likelihood in Three-way Contingency Tables," J.R. Statist. Soc. B 25, 220-233.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975) Discrete Multivariate Analysis, Cambridge, Mass.: M.I.T. Press

Darroch, J.N., Lauritzen, S.L., and Speed, T.P. (1980) "Markov Fields and Log-linear Interaction Models for Contingency Tables," Annals of Statistics, 8, 522-539.

Darroch, J.N., and Speed, T.P., (1983), "Additive and Multiplicative Models and Interactions," Ann. Statist., 11, 724-738.

Dawid, A.P. (1975), "On the Concepts of Sufficiency and Ancillarity in the Presence of Nuisance Parameters," Journal of the Royal Statistical Society, B, 37, 248-258.

Dawid, A.P. (1979a), "Conditional Independence in Statistical Theory," (with discussion), Journal of the Royal Statistical Society B 41, 1-31.

Dawid, A.P. (1979b), "Some Misleading Arguments Involving Condional Independence," Journal of the Royal Statistical Society B, 41, 249-252.

Dawid, A.P. (1980a), "A Bayesian Look at Nuisance Parameters," Bayesian Statistics: Proceedings of the First International Meeting held in Valencia (Spain), (Bernardo, et al, editors), University Press: Valencia.

Dawid, A.P. (1980b), "Conditional Independence for Statistical Operations," Annals of Statistics, 8, 598-617.

Fienberg, S.E. (1977), The Analysis of Cross Classified Categorical Data, Cambridge, Mass.: M.I.T. Press.

Goodman, L.A. (1970), "The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications," Journal of the American Statistical Association, 65, 226-256.

Goodman, L.A. (1971), "Partitioning of Chi-square, Analysis of Marginal Contingency Tables, and Estimation of Expected Frequencies in Multidimensional Contingency Tables," Journal of the American Statistical Association, 66, 339-344.

Haberman, S.J. (1974), <u>The Analysis of Frequency Data</u>, Chicago: University of Chicago Press.

Kiiveri, H., and Speed, T.P. (1982), "The Structural Analysis of Multivariate Data: a Review," Sociological Methodology, Samuel Leinhardt, Editor.

Kiiveri, H., Speed, T.P., and Carlin, J.B. (1984), "Recursive Causal Models," <u>Journal of the Australian Mathematical Society</u>, (Series A), 36, 30-52.

Koehn, U., and Thomas, D.L. (1975), "On Statistics Independent of a Sufficient Statistic: Basu's Lemma," <u>American Statistician</u>, 29, 40-42.

Lauritzen, S.L., Speed, T.P., and Vijayan, K. (1984), "Decomposable Graphs and Hypergraphs," <u>J. Austral. Math. Soc</u>. (Series A), 36, 12-29.

Lee, Yong Goo, (1986), "Independence Relationships for Multivariate Distributions," Ph.D. Thesis, University of Minnesota, School of Statistics.

Lee, Yong Goo, and Buehler, Robert J., (1986), "Independence Relationships for Multivariate Distributions," University of Minnesota, School of Statistics, Technical Report No. 464.

Speed, T.P., and Kiiveri, H.T. (1986), "Gaussian Markov Distributions over Finite Graphs," <u>Annals of Statistics</u> 14, 138-150.

Wermuth, N., (1976a), "Analogies between Multiplicative Models in Contingency Tables and Covariance Selection," Biometrics, 32, 95-108.

Wermuth, N., (1976b), "Model Search among Multiplicative Models," Biometrics, 32, 253-263.

Wermuth, N., (1980), "Linear Recursive Equations, Covariance Selection, and Path Analysis," Journal of the American Statistical Association, 75, 963-972.

Wermuth, N., and Lauritzen, S.L. (1983) "Graphical and Recursive Models for Contingency Tables," Biometrika, 70, 537-552.