Independence Relationships
for Multivariate Distributions

by

Yong Goo Lee[1] and
Robert J. Buehler

University of Minnesota
Technical Report No. 464
April 1986

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

---

[1] Partially supported by NIH Grant GM 25271.

I. Introduction

We are interested in independence conditions for multivariate distributions, their representation and interrelationships.

In working with contingency tables, Goodman (1970) p. 230, wrote $\bar{A} \otimes \bar{B}$ for A independent of B and $\bar{B} \otimes \bar{C}|\bar{A}$ for B independent of C given A. Darroch et al (1980) p. 533, use Goodman's notation.

In a general setting Dawid (1979a,b, 1980a,b) used $X \perp\!\!\!\perp Y$ to denote X independent of Y and $X \perp\!\!\!\perp Y|Z$ for X independent of Y for each fixed Z = z. Variations of Dawid's notation are found in Kiiveri and Speed (1982) p. 221, and Wermuth and Lauritzen (1983) p. 538. In view of Dawid's efforts to popularize his notation we will refer to it as Dawid notation even though it was anticipated by Goodman and no doubt others as well.

Goodman (1970) p. 234, wrote {AB}, {AC} for "fitted marginals," thereby indicating a hierarchical log linear contingency table model including terms u_{12}, u_{13} (in the notation of Bishop et al (1975)), but lacking u_{23} and u_{123}. Haberman (1974) used $\{\{1,2\}, \{1,3\}\}$ to represent the "generating class" of such models. Darroch et al (1980) p. 527, use Haberman's notation. The corresponding notation in Bishop et al (1975) p. 75, is C_{12}, C_{13} to denote a "sufficient configuration." Fienberg (1977) p. 40, preferred a slightly abbreviated "bracket notation" [12][13].

Kiiveri and Speed (1982) p. 215, and Kiiveri, Speed and Carlin (1984) p. 34, use abbreviations such as (1234) = (12)(3|2)(4|13) to denote a marginal-conditional factorization $p_{12}p_{3|2}p_{4|13}$. See also Geisser (1980) for a similar notation.

Models of these kinds are also related to the so-called graphical and decomposable models (see for example Darroch et al. (1980)).

In general terms we may think of first specifying a family of multivariate distributions, for example, nonsingular trivariate normal distributions or I by J by K contingency tables with all cells having positive probability. Then we introduce a classification scheme and ask how the family of distributions is partitioned by that method of classification. It is known for example that there are 128 log linear models describing three-way contingency tables of which 19 are hierarchical. The present report gives only an incomplete treatment of relationships among the following methods of classification: L = log linear, H = hierarchical, F = factorization, D = Dawid, MC = marginal-conditional, G = graphical, DC = decomposable.

In Section 2 we adopt a Goodman-Haberman-Fienberg notation, but instead of basing it on log linear models we take the broader view that it represents the factorization of a p.m.f. of p.d.f. (probability mass or density function). A basic result states that if π and π' are factorizations and C_π, $C_{\pi'}$ are the corresponding classes of functions, then a simple algorithm determines a third factorization π'' such that $C_\pi \cap C_{\pi'} = C_{\pi''}$.

In Section 3 we consider relationships between Dawid notation and factorization notation. It is shown that for three-way contingency tables with all cells having positive probability the Dawid notation partitions tables into 18 nonnull sets. Restriction to I by J by 2 tables reduces the count to 17.

Section 4 considers irregular cases where zero probabilities are allowed. A necessary and sufficient condition for $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ to imply $X \perp\!\!\!\perp (Y\ Z)$ is given. Failure of this condition has been noted earlier in connection with regularity conditions for a theorem of Basu on ancillary statistics.

Section 5 briefly describes relationsip of the present models to graphical and decomposable models.

2. Factorization Conditions

In this section we give a notation for factorization of a function $f(x_1, \ldots, x_N)$ and derive manipulative algorithms referred to as factorization calculus. In applications f will be either a p.m.f. or a p.d.f. (probability mass or density function).

Def. 2.1. The order of a model, N, is the number of arguments of $f(x_1, \ldots, x_N)$ (that is, the number of joint random variables).

Def. 2.2. A digit is a number 1,..., N.

Def. 2.3. A factor ϕ is a combination of distinct digits.

Following Fienberg (1977) we will write ϕ = [124] omitting commas. Haberman's notation would be {1, 2, 4}. The term factor is to be understood as a factor in the factorization of a density, and not as a factor in say a three-way layout (here identified with a digit).

Def. 2.4. A product is a set of factors (not necessarily distinct).

We write for example

(2.1) $$\pi = \{\phi_1\ \phi_2\ \phi_3\} = \{[1][12][23]\},$$

which is a mix of Haberman and Fienberg notation. Omission of commas follows Fienberg, while the outer braces (Haberman) identify a product as opposed to a factor. After reduction as defined below our product corresponds to Haberman's "generating class."

Def. 2.5. Reduction of a product means deletion of all factors which are proper subsets of other factors and deletion of all but one of any duplicated factors. 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, $\pi_1 \approx \pi_2$, if they reduce to the same minimal product.

Def. 2.8. The order of a factor is the number of digits: O[124] = 3.

Def. 2.9. The factor order of a product is the number of factors: O_f\{[12][23]\} = 2.

Def. 2.10. The vector order of a product gives the number of factors of different orders using exponents. VO{[1][23][24][25][34][4567]} = $1^1 2^4 4^1$.

Def. 2.11. A function $f(x_1, \ldots, x_N)$ has full support if $f>0$ for all

$x_1, \ldots, x_N$ in a domain S which is a direct product set: $S = S_1 \times \ldots \times S_N$ (where we think of S_j as the support of X_j).

Def. 2.12. The class C_π is the set of functions f which factor in accordance with π.

Example 2.1. If $\pi = \{[12][234][345]\}$, then $f \in C_\pi$ iff there exist a,b,c such that

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

Proposition 2.1. For the class of p.d.f.s. with full support on a finite set, the sets C_π for all minimal π correspond one-to-one with hierarchical log linear models whose generating classes (in the sense of Haberman (1974)) are π. Consequently the number of distinct factorizations equals the number of hierarchical models.

Def. 2.13. Set operations on factors:

$\phi \subseteq \phi'$ means every digit in ϕ is in ϕ'.

$\phi \cup \phi'$ means set of digits in either ϕ or ϕ'.

$\phi \cap \phi'$ means set of digits in both ϕ and ϕ'.

$\bigcup_{\phi\in\pi} \phi$ means set of all digits in some ϕ in π.

$\bigcap_{\phi\in\pi} \phi$ means set of all digits in every ϕ in π.

Def. 2.14. Set operations in products.

$\pi \subseteq \pi'$ means every factor in π is in π'.

$\pi \cup \pi'$ means set of factors in either π or π'.

$\pi \cap \pi'$ means set of factors in both π and π'.

$\pi \leq \pi'$ means every ϕ in π is a subset of some ϕ' in π'.

$\pi \wedge \pi'$ means the set of mn factors

$\phi_{ij} = \phi_i \cap \phi'_j$, $i = 1,\ldots, m$, $j = 1,\ldots, n$,

where $\pi = \{\phi_1 \ldots \phi_m\}$, $\pi' = \{\phi'_1 \ldots \phi'_n\}$.

It is trivial to show:

<u>Proposition 2.2</u>. $\pi \subseteq \pi'$ implies $\pi \leq \pi'$, and $\pi \leq \pi'$ implies $C_\pi \subseteq C_{\pi'}$.

<u>Example 2.2</u>. Take $N = 3$. Then $\{[1][2]\} \leq \{[12]\}$ expresses: If $f(x_1,x_2,x_3) = a(x_1)b(x_2)$ we can find $c(x_1,x_2)$ such that $f(x_1,x_2,x_3) = c(x_1,x_2)$.

There are a variety of ways of representing independence structures, and we are interested in studying the interrelationships among them. General types of representations include: (a) log linear models, (b) factorizations, (c) notations like $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$, (d) graphical method, (Darroch et al, 1980), (e) marginal and conditional factorizations. Not all conditions are represented by all methods. Goodman (1970) related (a,b,c) in certain cases.

<u>Proposition 2.3</u>. For $N = 3$ if $f(x,y,z)$ is a p.m.f. and $f \in C_\pi$, $\pi = \{[1][2]\}$, then $X \perp\!\!\!\perp Y$; but $X \perp\!\!\!\perp Y$ does not imply $f \in C_\pi$.

Thus the C_π classification of models may not do all that we might want.

For $N = 3$ let f_{12}, $f_{1.2}$, etc. denote marginal and conditional p.m.f.s so that $X \perp\!\!\!\perp Y|Z$ can be expressed as

$$f_{12.3} = f_{1.3}\, f_{2.3} \tag{2.3}$$

or equivalently

$$f_{123} = f_{13}f_{23}/f_3 = f_{13}f_{2.3} = f_{1.3}f_{23}, \tag{2.4}$$

from which $f = f_{123} \in C_\pi$, $\pi = \{[13][23]\}$. It is known that the converse is also true, according to:

Proposition 2.4. For $N = 3$ if $f(x,y,z)$ is a p.m.f. then $X \perp\!\!\!\perp Y|Z$ iff $f \in C_\pi$, $\pi = \{[13][23]\}$.

A more general version is:

Proposition 2.5. If ϕ_1, ϕ_2, ϕ_3 are a partition of digits 1,..., N, and X, Y, Z are corresponding vector variates, then $X \perp\!\!\!\perp Y|Z$ iff $f \in C_\pi$, $\pi = \{[\phi_1 \cup \phi_3][\phi_2 \cup \phi_3]\}$.

Proposition 2.6. For $N = 3$, $X \perp\!\!\!\perp (Y,Z)$ iff $f \in C_\pi$, $\pi = \{[1][23]\}$. For general N if X, Y, Z correspond to ϕ_1, ϕ_2, ϕ_3 as in Proposition 2.5, then $X \perp\!\!\!\perp (Y,Z)$ iff $f \in C_\pi$, $\pi = \{[\phi_1][\phi_2 \cup \phi_3]\}$.

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 $\pi = \{[13][23]\}$ write $C_\pi = C_{13/23}$, etc. Then in factorization notation the question translates to: Does $C_{13/23} \cap C_{12/23} = C_{1/23}$?

<u>Proposition 2.7</u>. For $N = 3$ for full support discrete p.m.f.s, $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ iff $X \perp\!\!\!\perp (Y,Z)$.

<u>Proof 1</u>. Log linear method. Let

$$\text{(2.5)} \qquad \log f = \mu + \alpha(x) + \beta(y) + \gamma(z) + (\alpha\beta)(x,y) + (\alpha\gamma)(x,z) + (\beta\gamma)(y,z) + (\alpha\beta\gamma)(x,y,z)$$

where $\alpha(x)$ etc. satisfy the usual analysis of variants constraints. Then $f \in C_{13/23}$ iff $(\alpha\beta) = (\alpha\beta\gamma) = 0$, $f \in C_{12/23}$ iff $(\alpha\gamma) = (\alpha\beta\gamma) = 0$, $f \in C_{1/23}$ iff $(\alpha\beta) = (\alpha\gamma) = (\alpha\beta\gamma) = 0$. The result follows.

<u>Proof 2</u>. Factorization method. Assume $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$. By (2.4)

$$\text{(2.6)} \qquad f = (f_{13}/f_3)f_{23} = (f_{12}/f_2)f_{23},$$

from which

$$\text{(2.7)} \qquad f_{23}\left(\frac{f_{13}}{f_3} - \frac{f_{12}}{f_2}\right) = 0$$

By the full support assumption $f_{23} \neq 0$, so that $f_{13}/f_3 = f_{12}/f_2$. 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.

<u>Proof 3</u>. Difference function method. We require a preliminary lemma.

<u>Lemma 2.1</u>. For N = 2 $f \in C_{1/2}$ iff $\Delta_2 \Delta_1 \log f = 0$ (where Δ_1, Δ_2 are difference operators), equality holding for all pairs of x values and pairs of y values in the support.

<u>Lemma 2.2</u>. For N = 3 $f \in C_{13/23}$ iff $\Delta_2 \Delta_1 \log f_{123} = 0$ for any z value and for any pairs of x and y values in the support.

<u>Proof</u>. Use Lemma 2.1 conditionally for each z.

<u>Continuation of Proof 3 of Proposition 2.7</u>. Let (c_1,c_2,c_3), (x,y,z) be two points in the support of f. Abbreviate $\eta(x,y,c_3) = -\log f(x,y,c_3)$ by $\eta_{12.}$ etc. If $X \perp\!\!\!\perp Y|Z$, then by Lemma 2.2

(2.8)
$$\eta_{12.} = \eta_{1..} + \eta_{.2.} - \eta_{...}$$
$$\eta_{123} = \eta_{1.3} + \eta_{.23} - \eta_{..3}$$

If $X \perp\!\!\!\perp Z|Y$ then

(2.9) $$\eta_{1.3} = \eta_{1..} + \eta_{..3} - \eta_{...}$$
$$\eta_{123} = \eta_{12.} + \eta_{.23} - \eta_{.2.}$$

Adding the four equations (2.8) and (2.9) gives

(2.10) $$\eta_{123} = \eta_{1..} + \eta_{.23} - \eta_{...}.$$

This expression shows $\eta(x,y,z)$ is a function of x plus a function of (y,z). Putting $f = e^{-\eta}$, an explicit solution is

(2.11) $$f(x,y,z) = a(x)\ b(y,z)$$

where

(2.12) $$a(x) = f(x,c_2,c_3),\ b(y,z) = f(c_1,y,z)/f(c_1,c_2,c_3),$$

showing that $f \in C_{1/23}$

Proof 4. Follows as a special case of factorization calculus, Proposition 2.8 below.

Def. 2.15. Let $c_1, \ldots, c_N$ be any fixed reference point in the domain of f. Then f_ϕ denotes the function

(2.13) $$f_\phi(x_1,\ldots, x_N) = f(y_1,\ldots, y_n) \text{ where } \begin{array}{l} y_i = x_i \text{ if } i \in \phi \\ y_i = c_i \text{ if } i \notin \phi. \end{array}$$

<u>Proposition 2.8</u>. Let $\phi_\Lambda = \phi_1 \cap \phi_2$ (Def. 2.13). If $f \in C_\pi$, $\pi = \{\phi_1\phi_2\}$ then

(2.14) $$f\, f_{\phi_\Lambda} = f_{\phi_1} f_{\phi_2}$$

and if $f_{\phi_\Lambda} \neq 0$ then an explicit factorization of f is

(2.15) $$f = \left(f_{\phi_1}/f_{\phi_\Lambda}\right) f_{\phi_2}.$$

<u>Proof</u>. There is no loss in generality in letting a single digit represent a set of digits. Digits not in $\phi_1 \cup \phi_2$ are irrelevant and can be dropped. Thus it suffices to take $\phi_1 = [12]$, $\phi_2 = [13]$, $\phi_\Lambda = [1]$. We find

(2.16) $$\begin{array}{lll} f & = f(x_1,x_2,x_3) & = a(x_1,x_2)b(x_1,x_3) \\ f_{\phi_\Lambda} & = f(x_1,c_2,c_3) & = a(x_1,c_2)b(x_1,c_3) \\ f_{\phi_1} & = f(x_1,x_2,c_3) & = a(x_1,x_2)b(x_1,c_3) \\ f_{\phi_2} & = f(x_1,c_2,x_3) & = a(x_1,c_2)b(x_1,x_3) \end{array}$$

which combine to give (2.15).

<u>Proposition 2.9</u>. If $f \in C_\pi$, $\pi = \{\phi_1\phi_2\phi_3\}$

then

(2.17) $$f \prod_{i \neq j} f_{\phi_{ij}} = f_{\phi_{123}} \prod_i f_{\phi_i}$$

where $\phi_{ij} = \phi_i \cap \phi_j$ and $\phi_{123} = \phi_1 \cap \phi_2 \cap \phi_3$.

Proof. Similar to preceding.

Corollary 2.9. If f has full support, f can be represented as a quotient of products in which each numerator factor and each denominator factor is obtained from f by setting certain variables equal to constants.

Proposition 2.10. If $f \in C_\pi$ has full support and if π has arbitary factor order n : $\pi = \{\phi_1, \ldots, \phi_n\}$, then an explicit factorization of f can be found as a quotient of two products in which each numerator factor and each denominator factor is obtained from f by putting certain variables equal to constant. That is all factors are of form f_ψ where ψ is expressible in terms of $\phi_1, \ldots, \phi_n$.

Proof. Proposition 2.9 generalizes to arbitrary factor order by union-intersection methods.

Proposition 2.11. For arbitrary ψ, π, if $f \in C_\pi$ then $f_\psi \in C_\pi$.

Proof. If f factors with arbitrary arguments then it factors with certain arguments equal to constants.

Proposition 2.12. Let $\pi = \{\phi_1\phi_2\}$, $\pi' = \{\phi_1'\phi_2'\}$. Then $f \in C_\pi \cap C_{\pi'}$ iff $f \in C_{\pi''}$, where $\pi'' = \pi\Lambda\pi'$ (see Def. 2.14).

Proof. "If." Follows from Proposition 2.2. "Only if." Factor f as in (2.15). By Proposition 2.11 all three quantities f_{ϕ_1}, f_{ϕ_2}, f_{ϕ_Λ} are in $G_{\pi'}$. Each can be explicitly factored according to Proposition 2.8 (replacing f by f_ϕ, π by π', etc.):

$$\begin{aligned} f_{\phi_1} &= \{f_{\phi_1\cap\phi_1'}/f_{\phi_1\cap\phi_\Lambda'}\}f_{\phi_1\cap\phi_2'} \\ f_{\phi_2} &= \{f_{\phi_2\cap\phi_1'}/f_{\phi_2\cap\phi_\Lambda'}\}f_{\phi_2\cap\phi_2'} \\ f_{\phi_\Lambda} &= \{f_{\phi_\Lambda\cap\phi_1'}/f_{\phi_\Lambda\cap\phi_\Lambda'}\}f_{\phi_\Lambda\cap\phi_2'} \end{aligned} \tag{2.18}$$

If we substitute (2.18) in (2.15) we get five factors in the numerator and four in the denominator. Each of these can be assigned to one or more of the four factors in the definition of π''. One explicit factorization is

$$\frac{f}{f_{\phi_\Lambda\cap\phi_\Lambda'}} = \frac{f_{\phi_1\cap\phi_1'}}{f_{\phi_1\cap\phi_\Lambda'}}\;\frac{f_{\phi_1\cap\phi_2'}}{f_{\phi_\Lambda\cap\phi_2'}}\;\frac{f_{\phi_2\cap\phi_1'}}{f_{\phi_\Lambda\cap\phi_1'}}\;\frac{f_{\phi_2\cap\phi_2'}}{f_{\phi_2\cap\phi_\Lambda'}} \tag{2.19}$$

Proposition 2.13 (factorization calculus). For any class of full support functions f, $C_\pi \cap C_{\pi'} = C_{\pi\Lambda\pi'}$ (see Def. 2.14 for $\pi \Lambda \pi'$.)

Proof. $f \in C_{\pi\Lambda\pi'}$ implies $f \in C_\pi$ and $f \in C_{\pi'}$ by Proposition 2.2. If $f \in C_\pi$ we can factor as described in Proposition 2.10. By Proposition 2.11 if $f \in C_{\pi'}$

then each of these factors can be further factored in accordance with π'. Each resulting factor will be contained in $\pi\wedge\pi'$.

Example 2.4. $\pi = \{[13][23]\}$, $\pi' = \{[12][23]\}$, $\pi\wedge\pi' = \{[1][3][2][23]\} = \{[1][23]\}$, which completes proof 4 of Proposition 2.7.

Example 2.5 (Markov chain). Assume $X_1 \perp\!\!\!\perp (X_3X_4)|X_2$ and $(X_1X_2) \perp\!\!\!\perp X_4|X_3$. By Proposition 2.4 $f \in C_\pi \cap C_{\pi'}$, $\pi = \{[12][234]\}$, $\pi' = \{[123][34]\}$, $\pi\wedge\pi' = \{[12][23][34]\}$. One explicit factorization is the marginal-conditional:

$$f_{1234} = f_{12}(f_{23}/f_2)(f_{34}/f_3) = f_{12}f_{3.2}f_{4.3}. \tag{2.20}$$

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 in the sense of Lehmann (1959), for example, 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 $C_{12/34/13/24}$ is evident. By Proposition 2.13 this is equivalent to $C_{124/134} \cap C_{123/234}$, which translates 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

(2.21)
$$\frac{f_{1234}}{f_{....}} = \frac{f_{12..}}{f_{1...}} \frac{f_{1.3.}}{f_{..3.}} \frac{f_{.2.4}}{f_{.2..}} \frac{f_{..34}}{f_{...4}} ,$$

where $f_{12..}(x_1,x_2,x_3,x_4) = f_{1234}(x_1,x_2c_3,c_4)$, etc. In contrast to Example 2.5 it appears to be impossible here to factor into marginal and conditional p.m.f.s.

3. Dawid Notation.

We have already mentioned Dawid's notation $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y|Z$, etc. A systematic treatment requires a definition of the scope of this notation. Let ϕ be a factor (Def. 2.3) and let X_ϕ be the subvector of $(X_1,\ldots, X_N)$ containing X_j for $j \in \phi$. We somewhat arbitrarily choose to allow the notation

(3.1)
$$X_{\phi_A} \perp\!\!\!\perp X_{\phi_B} | X_{\phi_C}$$

whenever ϕ_A, ϕ_B, ϕ_C are mutually disjoint and ϕ_A, ϕ_B are nonnull. Note that ϕ_A, ϕ_B disjoint rules out unneeded expressions like $X \perp\!\!\!\perp X|Y$ and ϕ_B, ϕ_C disjoint rules out $X \perp\!\!\!\perp (Y,Z)|Z$, which is no different from $X \perp\!\!\!\perp Y|Z$, and so unnecessary. We will understand that when ϕ_C is null, (3.1) means $X_{\phi_A} \perp\!\!\!\perp X_{\phi_B}$

When ϕ_A, ϕ_B, ϕ_C are exhaustive (contain all digits $1,\ldots, N$), then condition (3.1) is equivalent to $f \in C_\pi$, $\pi = \{\phi_1\phi_2\}$, $\phi_1 = \phi_A \cup \phi_C$, $\phi_2 = \phi_B \cup \phi_C$ (see Proposition 2.4).

If ϕ_A, ϕ_B, ϕ_C are not exhaustive then the Dawid condition (3.1) cannot be expressed in factorization notation. For example as was noted in Proposition 2.3, with $N = 3$, $X \perp\!\!\!\perp Y$ expresses $f_{12+} = f_{1++}f_{2++}$ where $f_{12+} = f_{12+}(x,y) =$

$\sum_z f_{123}(x,y,z)$, etc., and it is not possible to express the factorization of f_{12+} in terms of the factorization of f_{123}.

Conversely it is also impossible to express certain factorization conditions in Dawid notation. For N = 3 one example is $f \in C_\pi$, $\pi = \{[1][2]\}$, and another is $f \in C_\pi$, $\pi = \{[12][13][23]\}$.

Another choice of notation would allow expressions such as $\perp\!\!\!\perp (X,Y,Z)$, denoting mutual independence. At least in regular cases this mutual independence can be built up by combinations of marginal and conditional pairwise independence, so that nothing is actually lost by its omission.

For N = 3 we have nine different Dawid notations, three each from permutations of the three types: $X \perp\!\!\!\perp Y$, $X \perp\!\!\!\perp Y|Z$, $X \perp\!\!\!\perp (Y,Z)$.

For N = 4 we find seven types yielding a total of 55 permutations as follows: $X \perp\!\!\!\perp Y$ (6 permutations), $X \perp\!\!\!\perp (Y,Z)$ (12), $X \perp\!\!\!\perp Y|Z$ (9), $X \perp\!\!\!\perp (Y,Z)|W$ (12), $X \perp\!\!\!\perp Y|(Z,W)$ (6), $X \perp\!\!\!\perp (Y,Z,W)$ (4), $(X,Y) \perp\!\!\!\perp (Z,W)$ (6).

3.1. Dawid Partition for N = 3.

For N = 3 we have listed nine Dawid conditions, three each of three types. These have the potential of partitioning a family $\mathcal{F}$ of distributions into 2^9 subsets by considering all unions, intersections and complements. However, some cells will be empty, and our aim is to determine for a given $\mathcal{F}$ which cells are occupied. In this section we will show that if $\mathcal{F}$ is the family of full-support discrete distributions in $I \times J \times K$ points and we allow $I = 1,2,\ldots$, $J = 1,2,\ldots$, $K = 1,2,\ldots$, then 18 cells are occupied. If we restrict to $K = 2$, then 17 cells are occupied.

Proposition 3.1. For full support discrete models, $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ iff $X \perp\!\!\!\perp (Y,Z)$.

Proof. This is a restatement of Proposition 2.7.

Corollary. If we use the three conditions of type $X \perp\!\!\!\perp Y$ and the three conditions of type $X \perp\!\!\!\perp Y|Z$ then the three conditions of type $X \perp\!\!\!\perp (Y,Z)$ can be omitted without changing the partition of any full support discrete family $\mathcal{F}$.

Definition 3.1. With C, M, V, J denoting conditional, marginal, vector and joint we will use the abbreviated notation

C_x for $Y \perp\!\!\!\perp Z|X$

M_x for $Y \perp\!\!\!\perp Z$

V_x for $X \perp\!\!\!\perp (Y,Z)$

J for $\perp\!\!\!\perp (X,Y,Z)$

with C_y, etc., obtained by cyclic permutation. In this notation Proposition 3.1 reads $C_yC_z \Leftrightarrow V_x$.

Propositions 3.2 through 3.7, abbreviated below, are are easily proved and can be found in Goodman (1970).

Proposition 3.2. $C_yM_z \Leftrightarrow V_x$

3.3. $C_y C_z \iff V_x$

3.4. $V_x C_x \iff J$

3.5. $V_x M_x \iff J$

3.6. $V_x C_x \iff V_x$

3.7. $V_x M_z \iff V_x$

We will now consider the partition of the class of full-support discrete distributions into $64 = 2^6$ sets using the six Dawid notations

$X \perp\!\!\!\perp Y$ or M_z

$X \perp\!\!\!\perp Z$ or M_y

$Y \perp\!\!\!\perp Z$ or M_x

$X \perp\!\!\!\perp Y|Z$ or C_z

$X \perp\!\!\!\perp Z|Y$ or C_y

$Y \perp\!\!\!\perp Z|X$ or C_x

Each set will be represented in binary notation with for example (100010) denoting M_z and C_y true and the remaining four conditions false. The 64 conditions can be divided into 20 classes with 1, 3 or 6 members in each, where each class can be represented by a typical member and the others can be generated by cyclic permutations.

In Table 3.1 we list the groups and indicate which are possible. Each "possible" is justified by an example. Each "impossible" is justified by Proposition 3.1-3.7.

Proposition 3.8. There are 18 nonnull partition sets using Dawid notation with full-support discrete models.

Proof. In Table 3.1 groups 1,2,3,4,6,8,15,20 give 1+3+3+3+3+1+3+1 = 18 occupied cells.

Table 3.1
Possible and Impossible Combinations

Group	Binary representation	Number of permutations	Possible or Impossible	Justification
1	000000	1	P	Ex. 3.1
2	100000	3	P	Ex. 3.2
3	000100	3	P	Ex. 3.3
4	110000	3	P	Ex. 3.4
5	000110	3	I	Prop. 3.1
6	100100	3	P	Ex. 3.5
7	100010	6	I	Prop. 3.2
8	111000	1	P	Ex. 3.6
9	000111	1	I	Prop. 3.3,3.4
10	110100	6	I	Prop. 3.2,3.3,3.7
11	110001	3	I	Prop. 3.2,3.5
12	100110	6	I	Prop. 3.3,3.7
13	100011	3	I	Prop. 3.3,3.5
14	111100	3	I	Prop. 3.2,3.5
15	110110	3	P	Ex. 3.7
16	100111	3	I	Prop. 3.3,3.4
17	110101	6	I	Prop. 3.2,3.4,3.5
18	111110	3	I	Prop. 3.2,3.5
19	110111	3	I	Prop. 3.3,3.4
20	111111	1	P	Ex. 3.8

3.2. Examples of Discrete Distributions Satisfying Various Dawid Conditions.

Each example below represents a 2×2×2 discrete distribution (except 3.5 is 2×2×3). Probabilities are obtained by dividing by the sum over all cells. Each 2×2 table is a layer (z fixed) with x = row label and y = column label.

Example Number	Layer 1		Layer 2		Layer 3		Binary code
3.1	1	2	2	3			000000
	3	1	3	5			
3.2	1	3	2	3			100000
	1	2	1	2			
3.3	1	2	8	4			000100
	1	2	10	5			
3.4	3	5	1	3			110000
	4	8	2	4			
3.5	2	3	1	1	4	3	100100
	4	6	1	1	8	6	
3.6	1	3	3	5			111000
	3	5	5	11			
3.7	1	2	5	3			110110
	1	2	5	3			
3.8	1	1	1	1			111111
	1	1	1	1			

3.3. Some Examples of Distribution Families having Fewer Partition Sets.

Let F be the family of nonsingular trivariate normal distributions. The arguments showing which partition cells are empty in the discrete case can be repeated to give the same result. Alternatively independence relationships can be argued from zeros in the covariance matrix M and in $A = M^{-1}$.

Proposition 3.9. For the family of nonsingular trivariate normal distributions: (a) $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Z$, imply $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$. (b) $X \perp\!\!\!\perp Y$ and $X \perp\!\!\!\perp Y|Z$ imply either $X \perp\!\!\!\perp Z$ and $X \perp\!\!\!\perp Z|Y$, or $Y \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z|X$.

Proof. (a) Proof omitted. (b) Let M_{ij} = ij element of the covariance matrix (i = 1,2,3 j = 1,2,3), A_{ij} = ij element of M^{-1}. By hypothesis $M_{12} = 0$, $A_{12} = 0$. Representing A_{12} in terms of a cofactor gives $M_{13}M_{23} = 0$, which implies either X ⫫ Z or Y ⫫ Z.

Corollary. For the nonsingular trivariate normal family, the Dawid partition has 11 occupied cells.

Proof. Proposition 3.9 removes group 4 (code 110000) and group 6 (code 100100) with 3 members and group 8 (code 111000) with 1 member. Other possible cells in Table 3.1 remain possible, and are easily seen to correspond with zero patterns in M and A.

In the discrete case the partition may also depend on the size of the table. Example 3.5 is a 2×2×3 table because no 2×2×2 table would serve. This is shown in Proposition 3.10.

Proposition 3.10. Assume X,Y,Z has a full support discrete distribution where Z takes only two values, that is, the support is I×J×2. If X ⫫ Y and X ⫫ Y|Z then either X ⫫ Z or Y ⫫ Z.

Proof. Writing P(X=i,Y=j,Z=k) = p_{ijk}, we have from X ⫫ Y|Z,

(3.2) $$p_{ijk} = \lambda_k p_{ik} q_{jk}$$

where

(3.3) $$\sum_k \lambda_k = 1, \quad \sum_i p_{ik} = 1, \quad \sum_j q_{jk} = 1.$$

Define

(3.4) $$\tau_{ij} = P(X=i,Y=j) = \sum_k \lambda_k p_{ik} q_{jk}.$$

From X ⫫ Y we have for all pairs i,i',j,j',

(3.5) $$\tau_{ij}\tau_{i'j'} = \tau_{ij'}\tau_{i'j}$$

which is equivalent to

(3.6) $$(\lambda_1 p_{i1} q_{j1} + \lambda_2 p_{i2} q_{j2})(\lambda_1 p_{i'1} q_{j'1} + \lambda_2 p_{i'2} q_{j'2})$$
$$= (\lambda_1 p_{i1} q_{j'1} + \lambda_2 p_{i2} q_{j'2})(\lambda_1 p_{i'1} q_{j1} + \lambda_2 p_{i'2} q_{j2}).$$

The coefficients of λ_1^2 and λ_2^2 cancel, and the coefficients of $\lambda_1\lambda_2$ give

(3.7) $$p_{i1} q_{j1} p_{i'2} q_{j'2} + p_{i2} q_{j2} p_{i'1} q_{j'1}$$
$$= p_{i1} q_{j'1} p_{i'2} q_{j2} + p_{i2} q_{j'2} p_{i'1} q_{j1}$$

which can be factored as

(3.8) $$(p_{i1} p_{i'2} - p_{i2} p_{i'1})(q_{j1} q_{j'2} - q_{j2} q_{j'1}) = 0.$$

From (3.8) we have for all i,i',j,j'

$$\text{(3.9)} \qquad \text{either } \frac{p_{i'2}}{p_{i2}} = \frac{p_{i'1}}{p_{i1}} \quad \text{or} \quad \frac{q_{j'2}}{q_{j2}} = \frac{q_{j'1}}{q_{j1}}.$$

Call these two conditions $A_{ii'}$ and $B_{jj'}$. If $A_{ii'}$ holds for all i,i' then $X \perp\!\!\!\perp Z$. If $A_{ii'}$ fails for some i,i' then $B_{jj'}$ holds for all j,j' and $Y \perp\!\!\!\perp Z$.

Corollary. For I×J×2 tables the cell count in Table 3.1 is changed from 18 to 17.

Proof. Proposition 3.13 shows the cell 100100 is impossible. Because of the preferred Z direction in I×J×2, cells 010010 and 001001 are possible (Example 3.5), as are other possible cells in Table 3.1.

4. Irregular cases.

Example 2.3 shows how the factorization calculus fails when f does not have full support, that is, when there are zeros in the domain of f. In this section we will give necessary and sufficient conditions on the support of a discrete f for the factorization result to hold. Similar results have been given by Basu (1958), Koehn and Thomas (1975), Bishop et al (1975) Chapter 5, and Dawid (1979b, 1980a). 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 2.7. Fuller discussions can be found in the references cited above.

Let $f(x,y,z)$ be defined on a finite discrete set $S = S_x \times S_y \times S_z$, where S_x, S_y, S_z are the marginal supports: $f_1(x)>0$ for all $x \in S_x$ (where $f_1(x) = \Sigma_y \Sigma_z f(x,y,z)$), etc.).

Def. 4.1. Let S_{yz} be the marginal support of y and z:

$$S_{yz} = \{(y,z) | f_{23}(y,z)>0 \text{ where } f_{23}(y,z) = \Sigma_x f(x,y,z)\}$$

Two points (y,z) and (y',z') in S_{yz} are called y-linked if $y = y'$ and z-linked if $z = z'$. Two points are chain linked if they can be joined by a chain of y and z linked points.

Def. 4.2. 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 $(A \cap B) \cup (A^c \cap B^c)$. Then the set $A \times B$ will be called an x-oriented splitting set. (This terminology is adapted from Koehn and Thomas (1975).)

Proposition 4.1. Every pair of points in S_{yz} is chain linked iff there does not exist an x-oriented splitting set.

Proposition 4.2. 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.

Proof. Let x_o be any fixed point in S_x. Put

$$g_2(y) = f_{12}(x_0,y)/f_2(y), \quad g_3(z) = f_{13}(x_0,z)/f_3(z). \tag{4.1}$$

From (2.7) we have

$$f_{23}(y,z)\big(g_3(z) - g_2(y)\big) = 0 \quad \text{for all } (y,z) \in S_y \times S_z \tag{4.2}$$

For y-linked (y,z) and (y',z) in S_{yz}, $f_{23}(y,z) > 0$, $f_{23}(y',z) > 0$, giving

$$g_3(z) = g_2(y) \quad \text{and} \quad g_3(z) = g_2(y') \tag{4.3}$$

so that $g_2(y) = g_2(y') = g_3(z)$. Similarly for z-linked points $g_3(z) = g_3(z') = g_2(y)$. Thus for fixed x_0, $g_2(y)$ and $g_3(z)$ take constant values in any chain linked set. The result follows.

Proposition 4.3. $X \perp\!\!\!\perp Y|Z$ and $X \perp\!\!\!\perp Z|Y$ imply $X \perp\!\!\!\perp (Y,Z)$ iff there does not exist an x-oriented splitting set.

Proof. If there does not exist a splitting set then Proposition 4.1 shows that all points are chain linked and Proposition 4.2 shows that $f_{13}(x,z)/f_3(z)$

depends only on x, and can be called a(x). From (2.6), $f(x,y,z) = a(x)f_{23}(y,z)$, showing $X \perp\!\!\!\perp (Y,Z)$. Example 2.3 shows $X \perp\!\!\!\perp (Y,Z)$ can fail when there is a splitting set.

5. Graphical and Decomposable Models.

Factorization models are one-to-one with hierarchical models (Proposition 2.1). A subclass of hierarchical models are the decomposable models of Goodman 1970) and Haberman (1974). Intermediate between hierarchical and decomposable models are the graphical models of Darroch et al (1980).

Let us associate with each digit a vertex. Given any product π, construct a graph as follows: two vertices are joined with an (undirected) line iff the corresponding digits occur together in any factor. 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.

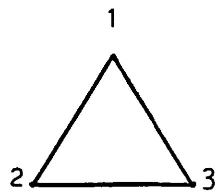
By the above construction any product π determines a graph and that graph determines (and is determined by) its set of cliques. Can the cliques be used to recover the product π? Sometimes, but not always.

Definition 5.1. A minimal product π is graphical if the cliques it defines are the same as its factors.

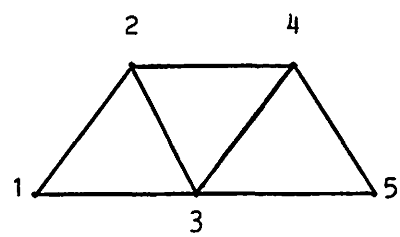
Example 5.1. Consider the products

(5.1) $$\pi_1 = \{[123]\},\ \pi_2 = \{[12][13][23]\},\ \pi_3 = \{[123][234][345]\}.$$

Both π_1 and π_2 have the graph



and a single clique [123], corresponding to the single factor in π_1, but different from π_2. Thus π_1 is graphical but π_2 is not. The graph of π_3 is



from which it is seen that the cliques are [123], [234], [345], the same as the factors, so that π_3 is graphical.

Among models of order 5, product π_3 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.

<u>Definition 5.2</u>. A minimal product π is <u>decomposable</u> if $\pi = \pi_1 \cup \pi_2$ (this and the following notation was given in Def. 2.13 and 2.14) where $\pi_1 \cap \pi_2$ = null

set of factors, and there exist $\phi_1 \in \pi_1$, $\phi_2 \in \pi_2$ such that

$$(5.2) \qquad \left(\bigcup_{\phi\in\pi_1} \phi\right) \cap \left(\bigcup_{\phi\in\pi_2} \phi\right) = \phi_1 \cap \phi_2.$$

This definition corresponds to Haberman's (1974) p. 166, decomposable generating class. The definition of Darroch et al (1980) p. 522, of decomposable model agrees with Haberman's decomposable generating class.

Decomposability of a generating class is necessary and sufficient for existence of closed form maximum likelihood estimators in hierarchical contingency table models (Goodman (1970, 1971), Haberman (1970, 1974)).

The analogous result for products π is that π can be factored in accordance with its factors with each factor representing either a marginal or a conditional distribution iff π is decomposable.

Example 5.2. The product $\pi_4 = \{[12][23]\}$ is decomposable. There are two ways to factor into marginal and conditional factors:

$$(5.3) \qquad f_{123} = f_{12}(f_{23}/f_2) = f_{12}f_{3.2} = (f_{12}/f_2)f_{23} = f_{1.2}f_{23}$$

In general any product having only two factors can be factored analogously.

Example 5.3. If $\pi_3 = \{[123][234][345]\}$ as in Example 5.1, there are several ways to factor one of which is

(5.4) $$f_{12345} = f_{123}f_{4.23}f_{5.34} = f_{123}(f_{234}/f_{23})(f_{345}/f_{34}).$$

Example 5.4. Consider again the exponential family model of Example 2.6 in which $f_{1234} \varepsilon C_{\pi_5}$ with $\pi_5 = \{[12][34][13][24]\}$. This π_5 is nondecomposable according to Haberman's criterion (Definition 5.2) and it is impossible to factor f_{1234} into four factors with each factor either a marginal or a contiional distribution. The graph for this model is

which is the simplest nondecomposable model in the catalog of Darroch et al (1980) p 536. The three permutations of π_5 are the only three nondecomposable models of dimension 4 out of 113 graphical models and 167 hierarchical models.

REFERENCES

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

Basu, D. (1977), "On the Elimination of Nuisance Parameters," Journal of the American Statistical Association, 72, 355-366.

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

Barndorff-Nielsen, O. (1979), Information and Eponential Families in Statistical Theory, New York: John Wiley.

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.

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 Conditional 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.

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

Geisser, S. (1979), Discussion of Dawid (1979a).

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. (1970), The General Log-Linear Model, Ph.D. Thesis, Department of Statistics, University of Chicago.

Haberman, S.J. (1974), The Analysis of Frequency Data, 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," Journal of the Australian Mathematical Society, (Series A), 36, 30-52.

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

Lehmann, E., (1959), Testing Statistical Hypotheses, New York: John Wiley.

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