<**Research Paper**>

Journal of the Korean Statistical Society
Vol. 19, No. 1, 1990

Factorization Models and Other Representation of Independence +

Yong-Goo Lee*

ABSTRACT

Factorization models are a generalization of hierarchical loglinear models which apply equally to discrete and continuous distributions. In regular(strictly positive) cases the intersection 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.

1. Introduction

The purpose of this paper is to review methods of representing independence in multivariate distributions and to relate them to factorization models, a generalization of hierarchical log linear models based on factorization of density functions. Factorization models appear more or less explicitly for example in Darroch et al.(1980), Darroch and Speed(1983), and Kiiveri et al.(1984) and more implicitly in work such as Wermuth(1976a), (1980), David(1979a), (1980), and Lauritzen et al.(1984).

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.

+ This research was partially supported by the Chung Ang University Research Fund
* Department of Applied Statistics, Chung Ang University, Seoul , Korea

The way in which factorization models generalize hierarchical log linear models can be seen by an example. A probability mass function (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} (Goodeman(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(1976a, b) to 12/13.

2. Definitions for Factorization Models

In this section, we will define some notations which is useful for understanding factorization models and its manipulation.

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

Definition 2.2 A *factor* is a number 1,⋯,N. The set {1,⋯,N} will be denoted by $\mathcal{N}$. This terminology agrees with Darroch et al.(1980).

Definition 2.3 Any Subset of $\mathcal{N}$, say $a \subseteq \mathcal{N}$, will be called a *term*.

Definition 2.4 A *product* $\mathcal{A}$ is a set of terms(not necessarily distinct). we will variously write for example.

$$\mathcal{A} = \{a, b, c\} = \{[1], [12], [23]\} = 1/12/23, \tag{2.1}$$

which is a mix of Haberman, Fienberg and Wermutch 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 a function f into a product.

Definition 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]}.

Definition 2.6 A product is *minimal* if it has no reduction.

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

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

It may be helpful to think of $f \in C_{\mathcal{A}}$ as equivalent to log f belonging to a linear subspace. In notation close to that of Darroch, et al.(1980) p.54, and Darroch and Speed(1983) p.725, $f \in C_{\mathcal{A}}$ iff

$$\log f = \sum_{a \in \mathcal{A}} \xi_a(X_a), \tag{2.2}$$

where X_a is the set of X_i for $i \in a$

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

$$f(X_1, \cdots, X_5) = a(X_1, X_2) b(X_2, X_3, X_4) c(X_3, X_4, X_5).$$

Definition 2.9 Set operations on products.

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

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

$\mathcal{A} < \mathcal{B}$ means $\mathcal{A} \leq \mathcal{B}$ but $\mathcal{A} \neq \mathcal{B}$.

$\mathcal{A} \wedge \mathcal{B}$ means the set of mn terms

$$C_{ij} = a_i \cap b_j,\ i = 1, \cdots, m,\ j = 1, \cdots, n, \text{ where}$$
$$\mathcal{A} = \{a_1, \cdots, a_m\},\ \mathcal{B} = \{b_1, \cdots, b_n\}.$$

The notation $\mathcal{A} \Lambda \mathcal{B}$ agress with Lauritzen et al.(1984), P.16.

Proposition 2.1 $\mathcal{A} \subseteq \mathcal{B}$ implies $\mathcal{A} \leq \mathcal{B}$, and $\mathcal{A} \leq \mathcal{B}$ implies $C_{\mathcal{A}} \subseteq C_{\mathcal{B}}$.

Proof. First part is a result of the definition 2.9.

Let us define that $\mathcal{A} = \{a_1, \cdots, a_m\}$ and $\mathcal{B} = \{b_1, \cdots, b_n\}$.

If $f \in C_{\mathcal{A}}$, then f can be factored in accordance with $\mathcal{A}$. And by the given condition, every a_i, $i = 1, \cdots, m$ is a subset of some b_j, $j = 1, \cdots, n$ which implies that f can also be factored in accordance with $\mathcal{B}$. This implies that $f \in C_{\mathcal{B}}$. The result follows.

The following example will be helpful to understand the proposition 2.1

Example 2.2 Take N=3. Let $\mathcal{A} = \{[1], [2], [3]\}$ and $\mathcal{B} = \{[1], [23]\}$, then $\mathcal{A} \leq \mathcal{B}$ and $f \in C_{\mathcal{A}}$ if there exists a, b, c such that

$f = a(X_1)b(X_2)c(X_3)$. Assume that $b(X_2)c(X_3) = d(X_2, X_3)$, then

$f = a(X_1)d(X_1, X_3)$, which implies that $f \in C_{\mathcal{B}}$.

3. Factorizations Calculus

The distributions can be either discrete or continuous, and our treatment is non-measure theoretic and assumes conditional densities to be defined as the quotient of joint and marginal densities.

Using his notation for conditional independence, Dawid(1979a) pointed out that

$$X \perp\!\!\!\perp Y \mid Z \text{ iff } f(X, Y, Z) = a(X, Z)b(Y, Z). \tag{3.1}$$

Example 3.1 If $f(X_1,\cdots,X_6)=ce^{-Q/2}$, $Q=\Sigma\Sigma a_{ij}X_iX_j$, $a_{13}=a_{14}=a_{23}=a_{24}=0$, then $(X_1, X_2) \perp\!\!\!\perp (X_3, X_4) \mid (X_5, X_6)$. 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 (3.1) establishes the result by factorization, avoiding matrices.

A more general version of (3.1) is:

Proposition 3.1 (The independence-factorization connection)

If a, b, c are a partition of $\mathcal{N}=\{1,\cdots,N\}$ and X, Y, Z are corresponding vector variates then

$$X \perp\!\!\!\perp Y \mid Z \quad \text{iff } f \in C_{\mathcal{A}}, \mathcal{A}=\{a\cup c, b\cup c\}. \tag{3.2}$$

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

Example 3.2 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 $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ but $X \perp\!\!\!\perp (Y, Z)$ is false.

When do $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ imply $X \perp\!\!\!\perp (Y, Z)$?

For $\mathcal{A}=\{[12], [23]\}$ write $C_{\mathcal{A}}=C_{12/23}$, etc. Then in factorization notation the question translates to:

Does $C_{13/23} \cap C_{12/23} = C_{1/23}$? Dawid(1980), Sec. 6 and 7, gives a measure theoretic treatment.

Proposition 3.2 For N=3 if f is strictly positive, then $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ iff $X \perp\!\!\!\perp (Y, Z)$.

Proof. The proof of "if" part is trivial and will be omitted here. Assume $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$, then we can write

$$f=(f_{13}/f_3)f_{23}=(f_{12}/f_2)f_{23}, \tag{3.3}$$

from which

$$f_{23}(f_{13}/f_3-f_{12}/f_2)=0. \tag{3.4}$$

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.

For routine manipulations the following generalization of Proposition 3.2 is useful.

Proposition 3.3 (Factorization Calculus) For any class of strictly positive functions f, $C_{\mathcal{A}} \cap C_{\mathcal{B}} = C_{\mathcal{A}\wedge\mathcal{B}}$.

Proof. By definition, $\mathcal{A}\Lambda\mathcal{B} \leq \mathcal{A}$ and $\mathcal{A}\Lambda\mathcal{B} \leq \mathcal{B}$. By Proposition 2.2, $C_{\mathcal{A}\wedge\mathcal{B}} \subseteq C_{\mathcal{A}}$ and $C_{\mathcal{A}\wedge\mathcal{B}} \leq C_{\mathcal{B}}$, and so $C_{\mathcal{A}\cap\mathcal{B}} \subseteq C_{\mathcal{A}} \cap C_{\mathcal{B}}$. Assume that $f \in C_{\mathcal{A}} \cap C_{\mathcal{B}}$. Then there exist $a_1,\cdots,a_m$ and $b_1,\cdots,b_n$ such that $\mathcal{A}=\{a_1,\cdots,a_m\}$ and $\mathcal{B}=\{b_1,\cdots,b_n\}$ and f is factored in accordance with $\mathcal{A}$ and $\mathcal{B}$. This implies that f is factored in accordance with $\mathcal{C}=\mathcal{A}\Lambda\mathcal{B}=\{a_i \cap b_j\}$, $i=1,\cdots,m$, $j=1,\cdots,n$ which implies $f \in C_{\mathcal{C}}$, so $C_{\mathcal{A}} \cap C_{\mathcal{B}} \subseteq C_{\mathcal{A}\wedge\mathcal{B}}$.

Example 3.3 By proposition 3.1, $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ translate to $\mathcal{A}=13/23$ and $\mathcal{B}=12/23$, giving $A\Lambda B=1/3/2/23 \approx 1/23$, which translates to $X \perp\!\!\!\perp (Y, Z)$, showing Proposition 3.2 to be a special case of Proposition 3.3.

Example 3.4 (Markov chain) Assuming $X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2$ and $(X_1, X_2) \perp\!\!\!\perp X_4 \mid X_3$ gives $\mathcal{A}=12/234$, $\mathcal{B}=123/34$, $\mathcal{A}\Lambda\mathcal{B}=12/23/34$. One explicit factorization is the marginal-conditional:

$$f_{1234}=f_{12}(f_{23}/f_2)(f_{34}/f_3)=f_{12}f_{3.2}f_{4.2}. \tag{3.5}$$

Example 3.5 (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 image 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 3.3 this is equivalent to $C_{124/134} \cap C_{123/234}$, which translated by Proposition 3.1 to $X \perp\!\!\!\perp \beta \mid (Y, \alpha)$ and $Y \perp\!\!\!\perp \alpha \mid (X, \beta)$.

In the terminology of Dawid(1975), Basu(1977) and Barndorff-Nielsen(1978), X and Y respectively are specific sufficient for α and β and specific ancillary for β and α. In constrast to Example 3.4 it is impossible here to factor into marginal and conditional p.m.f.'s.

Example 3.6 (ZPA conditions) Take $N=5$.

Wermuth(1976a) write 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) P254-5, are a special case of the present factorization calculus. To see this, apply Proposition 3.3 to wermuth's example of finding the conjunction of ZPA(1, 2), ZPA(1, 3) and ZPA(2, 3). We find (1345/2345)Λ(1245/2345)=145/2345, and (145/2345)Λ(1345/1245)=145/245/345.

4. Irregular Cases

Example 3.2 shows how the factorization calculus fails 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 certain factorization results to hold. Similar results have been given by Basu(1958), Koehn and Thomas(1975), Bishop et al.(1975) Chapter 5, Dawid(1979b, 1980).

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 $U \perp\!\!\!\perp \theta \mid T$, I=independence condition expressed as $U \perp\!\!\!\perp T \mid \gamma$, A=ancillary condition expressed as $U \perp\!\!\!\perp \theta$. Basu's Theorem 2 states: S and I imply A, which follows from Proposition 3.2. Further 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 *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.

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

Starting with equation(3.3) we can show:

Proposition 4.2 Assume $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$. 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 x only.

Proof. Let x_0 be any fixed point in S_X. Put

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

From (3.4) we have

$$f_{23}(y, z)(g_3(z) - g_2(y)) = 0 \text{ for all } (y, z) \in S_Y x S_z. \tag{4.2}$$

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

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

So that $g_2(y) = g_2(y') = g_3(z)$. Similarly for y-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, and the result follows.

Proposition 4.3 $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ 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 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). Thus $f(x, y, z) = a(x)\, f_{23}(y, z)$, showing $X \perp\!\!\!\perp (Y, Z)$.

Example 3.2 is an example that if there exists a yz-splitting set then $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ does not imply $X \perp\!\!\!\perp (Y, Z)$. This counter example implies that if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp Z \mid Y$ imply $X \perp\!\!\!\perp (Y, Z)$ then there does not exist a yz-splitting set.

Proposition 4.3 shows that without the full support assumption the equality in Proposition 3.2 holds iff there does not exist a yz-splitting set. And this result can be applied for justifying Basu's Theorem 2.

Acknowledgement

I wish to thank to the editor and the anonymous referees for their helpful comments and suggestions. Their detailed comments were very useful in improving quality of the paper.

References

1. Barndorff-Nielson, O.(1978). *Information and Exponential Families in Statistical Theory,* New York: John Wiley.
2. Basu, D.(1958). On Statistics Independent of Sufficient Statistics, *Sankhya,* Vol. 20, 223-226.
3. Basu, D.(1982). Basu Theorems, *Encyclopedia of Statistical Sciences,* Kotz, Johnson and Read, Editors, Vol. 1, 193-196, New York: John Wiley.
4. Bishop, Y.M.M., Fienberg, S.E. and Holland, P.W.(1975). *Discrete Multivariate Analysis,* Cambridge, Mas.: MIT Press.
5. Darroch, J.N., Lauritzen, S.L., and Speed, T.P.(1980). Markov Fields and Log-linear Interaction Models for Contingency Tables, *The Annals of Statistics,* Vol. 8, 522-539.
6. Darroch, J.N. and Speed, T.P.(1983). Additive and Multiplicative Models and Interactions, *The Annals of Statistics,* Vol. 11, 724-738.
7. Dawid, A.P.(1975). On the Concepts of Sufficiency and Ancillarity in the Presence of Nuisance Parameters, *Journal of the Royal Statistical Society,* B, Vol. 37, 248-258.
8. Dawid, A.P.(1979a). Conditional Independence in Statistical Theory, (with discussion), *Journal of the Royal Statistical Society,* B, Vol. 41, 1-31.
9. Dawid, A.P.(1979b). Some Misleading Arguments Involving Conditional Independence, *Journal of the Royal Statistical Society,* B, Vol. 41, 249-252.
10. Dawid, A.P.(1980). Conditional Independence for Statistical Operations, *Annals of Statistics,* Vol. 8, 598-617.
11. Fienberg, S.E.(1977). *The Analysis of Cross Classified Categorical Data,* Cambridge, Mass.: MIT Press.
12. Goodman, L.A.(1970). The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classifications, *Journal of the American Statistical Associations,* Vol. 65, 226-256.
13. Haberman, S.J.(1974). *The Analysis of Frequency Data,* Chicago: University of Chicago Press.
14. Kiiveri, H., Speed, T.P., and Carlin, J.B.(1984). Recursive Causal Models, *Journal of the*

Australian Mathematical Society, (Series A), Vol. 36, 30-52.

15. Koehn, U. and Thomas, D.L.(1975). On Statistics Independent of a Sufficient Statistic: Basu's Lemma, *American Statistician,* Vol. 29, 40-42.
16. Lauritzen, S.L., Speed, T.P. and Vijayan, K.(1984). Decomposable Graphs and Hypergraphs, *Journal of the Australian Mathematical Society,* (Series A), Vol. 36, 12-29.
17. Wermuth, N.(1976a). Analogies between Multiplicative Models in Contingency Tables and Covariance Selection, *Biometrics,* Vol. 32, 95-108.
18. Wermuth, N.(1975b). Model Search among Multiplicative Models, *Biometrics,* Vol. 32, 253-263.
19. Wermuth, N.(1980). Linear Recursive Equations, Covariance Selection, and Path Analysis, *Journal of the American Statistical Association,* Vol. 75, 963-972.