Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT•

Journal of Multivariate Analysis 84 (2003) 403–409

http://www.elsevier.com/locate/jmva

Some equivalence results concerning multiplicative lattice decompositions of multivariate densities

Edward H. Ip,[a,*,1] Yuchung J. Wang,[b] and Yeong-nan Yeh[c,2]

[a] *Information and Operations Management Department, Marshall School of Business, University of Southern California, 3670 Trousdale Pkwy Bridge Hall, 401R, Los Angeles, California, 90089, USA*
[b] *Department of Mathematical Sciences, Rutgers University, USA*
[c] *Institute of Mathematics, Academia Sinica, Taiwan*

Received 30 January 2001

Abstract

This note provides some equivalence results across the partition lattice, the monotypic lattice, and the subset lattice, for decomposing a multivariate density.

AMS 1991 subject classifications: 62E10; 62H20

Keywords: Lancaster model; Bahadur model; Partition lattice; Combinatorics

1. Introduction

Lattice decomposition of a joint density is an approach that uses modern algebra to delineate the structure of a multivariate density. Streitberg [12,13] used the partition lattice to decompose a density in order to obtain measures of multivariate dependency. Ip et al. [9] showed that the Lancaster/Bahadur decomposition [2,10] can also be expressed as a lattice decomposition similar to Streitberg's. A characterizing property of both the Streitberg and Lancaster/Bahadur lattice

*Corresponding author.
E-mail addresses: eddie.ip@marshall.usc.edu (E.H. Ip), mayeh@ccvax.sinica.tw (Y.-n. Yeh).
[1] Partially supported by a research Grant from the Marshall School of Business, University of Southern California.
[2] Partially supported by the Grant NSC89-2115-M001-02 from the National Science Council, Taiwan.

0047-259X/03/$ - see front matter
doi:10.1016/S0047-259X(02)00051-9

decompositions is the reproducibility of the resulting interactions [11]. In other words, the decompositions lead to parameters with meanings that are invariant across different marginal distributions. This is a highly desirable property for marginal inference in categorical data analysis [7]. Unfortunately, this property is not present in the method most commonly used to analyze categorical data, which is the loglinear model [3,8]. In contrast to the multiplicative loglinear model, in which probabilities are decomposed on the log-scale, the interactions in both the Streitberg and the Lancaster/Bahadur models are additive.

In this article, we provide some equivalence results concerning the representation of a multivariate joint density across several lattices, including those described above. The equivalence representation is both multiplicative and reproducible. The results are important for several reasons. First, although multiplicative models are now routinely used in statistics, most models are not reproducible [7]. Second, the equivalence decomposition forms the basis for defining a new class of reproducible measures of association. Finally, the combinatoric methods used in our proof are general and can be applied to functions other than density [9].

2. Background

2.1. Lattice preliminary

Let $N = \{1, \dots, n\}$. A partition σ of N is a collection of disjoint non-empty subsets $\sigma_1, \dots, \sigma_k$, $\sigma_i \subset N, i = 1, \dots, k$, where $\bigcup_i \sigma_i = N$. Write $\sigma = \sigma_1|\cdots|\sigma_k$. Each σ_i is referred to as a block of σ. For example, the partition 12|3|4 consists of the blocks 12, 3, and 4. A partial order $\prec$ is defined by refinement: for any partitions σ, π of N, $\sigma \prec \pi$ if each block of σ is contained in a block of π. For any two partitions σ and π, there exist a unique least upper bound $\sigma \vee \pi$ and a unique greatest lower bound $\sigma \wedge \pi$. For instance, $12|3|4 \vee 1|2|34 = 12|34$ and $1|234 \wedge 12|34 = 1|2|34$. Thus the collection $S = P(N)$ of all partitions, together with the partial order $\prec$, form a lattice $\mathscr{L}(S) = (S, \prec)$.

The subset lattice $S = B(N)$ is formed by the power set 2^N, together with the partial ordering $\prec$ defined by inclusion: $A, B \in B(N)$, $A \prec B$ if and only if $A \subset B$, and $A \vee B = A \cup B$ and $A \vee B = A \cap B$.

For any finite lattice $\mathscr{L}(S)$, there exists supremum $\hat{1}$ and infimum $\hat{0}$. As an example, for the partition lattice, $\hat{1} = 1\cdots n$, $\hat{0} = 1|2|\cdots|n$; for the subset lattice, $\hat{1} = N$, $\hat{0} = \emptyset$.

Let S denote a finite set and $\prec$ be a partial order relationship defined on S. For $\tau \in S$, $\triangle(\tau)$ denotes a real-valued function defined on S. The *sum function* f of $\triangle$ at σ is given by

$$f(\sigma) = \sum_{\tau \prec \sigma} \triangle(\tau). \tag{1}$$

The calculus of inversion on $\mathscr{L}(S)$ is determined by the Möbius inversion function. Specifically,

$$\triangle(\sigma) = \sum_{\tau \prec \sigma} \mu(\tau, \sigma) f(\tau). \tag{2}$$

Alternatively, we can construct an operator $\triangle_\sigma$ for each $\sigma \in \mathscr{L}(S)$ and write (2) in shorthand as $\triangle_\sigma f = \sum \mu(\tau, \sigma) f_\tau$. We also define $\triangle f := \triangle_{\hat{1}} f$.

A general reference for lattice theory is Aigner [1].

2.2. Three lattice decompositions

Let F denote a distribution function. For $\pi = \pi_1 | \cdots | \pi_k \in P(N)$, define $\triangle_\pi^{(P)} F = \prod \triangle_{\pi_i}^{(P)} F_{\pi_i}$, where F_{π_i} is the marginal distribution formed by the variables that belong to the non-empty subset π_i of N. The following description about operators are rather general and they work on densities the same way they work on measures. We introduce three different lattices for decomposing a joint multivariate density: the monotypic lattice, the partition lattice, and the subset lattice.

The monotypic lattice representation describes the decomposition of a multivariate density due to the Lancaster/Bahadur symbolic notation [10]:

$$\triangle F = \prod_i (F_i^\star - F_i), \tag{3}$$

where in the expansion $F_{i_1 \cdots i_k}$, the marginal distribution of $(X_{i_1}, \ldots, X_{i_k})$, is understood to substitute the product $F_{i_1}^\star \cdots F_{i_k}^\star$. For example, the term $F_1^\star F_2^\star F_3$ in the expansion $(F_1^\star - F_1)(F_2^\star - F_2)(F_3^\star - F_3)$ is substituted by $F_{12}F_3$. Ip et al. [9] showed that the symbolic representation (3) can be expressed in terms of interaction on a monotypic lattice $M(N)$, which is formed by removing from a partition lattice elements that contain two or more non-singleton blocks. Fig. 1 shows the structure of the monotypic lattice $M(N)$ for $n = 4$. The Lancaster/Bahadur interaction for $n = 4$ is

$$\begin{aligned}\triangle^{(M)} F = & F_{1234} - (F_{234|1} + F_{134|2} + F_{124|3} + F_{123|4}) \\ & + (F_{1|2|34} + F_{14|2|3} + F_{1|3|24} \\ & + F_{13|2|4} + F_{23|1|4} + F_{12|3|4}) - 3F_{1|2|3|4},\end{aligned} \tag{4}$$

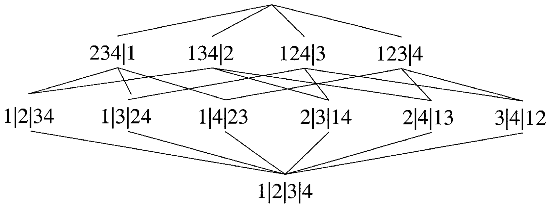

Fig. 1. Monotypic lattice with $n = 4$.

which includes only terms with at most one non-singleton block (terms of the type $F_{12}F_{34}$ are missing).

The Möbius function μ on $\mathscr{L}(M(N))$ is given by

$$\mu^{(M)}(\sigma, \hat{1}) = \begin{cases} (-1)^{n-1}(n-1) & \text{if } \sigma = \hat{0}, \\ (-1)^{|\sigma|-1} & \text{otherwise,} \end{cases}$$

where $|\sigma|$ is the number of blocks of σ [9].

Streitberg's interaction is motivated by a desire to correct a deficiency in the Lancaster/Bahadur decomposition, namely, the fact that the latter does not satisfy the interaction axiom. The axiom states that if the joint density factorizes, then the highest-order interaction vanishes—i.e., $\triangle_{\hat{1}} F = 0$. The Lancaster/Bahadur decomposition fails the interaction axiom for $n \geqslant 4$. For example, when $F_{1234} = F_{12}F_{34}$, $\triangle^{(M)} F = (F_{12} - F_1F_2)(F_{34} - F_3F_4)$, which in general, is not identically zero. To correct for the deficiency, Streitberg [12] used instead a partition lattice to decompose the joint density. The Möbius inversion function of the partition lattice $\mathscr{L}(P(N))$ is well known [1]:

$$\mu^{(P)}(\sigma, \hat{1}) = (-1)^{|\sigma|-1}(|\sigma| - 1)!.$$

As an example, when $n = 4$, Streitberg's measure of interaction is

$$\begin{aligned} \triangle^{(P)} F = {} & F_{1234} - (F_{123|4} + F_{124|3} + F_{134|2} + F_{234|1}) \\ & - (F_{12|34} + F_{13|24} + F_{14|23}) + 2(F_{1|2|34} + F_{1|3|24} + F_{1|4|23} \\ & + F_{2|3|14} + F_{2|4|13} + F_{3|4|12}) - 6F_{1|2|3|4}. \end{aligned} \tag{5}$$

The third method for decomposing the joint density is over the subset lattice $B(N)$. By convention, we set $F_{\emptyset} = 0$. The Möbius inversion function of the subset lattice is given by

$$\mu^{(B)}(\sigma, \hat{1}) = (-1)^{n-||\sigma||},$$

where $||\sigma||$ is the number of elements in σ. For example, when $n = 4$,

$$\begin{aligned} \triangle^{(B)} F = {} & F_{1234} - (F_{134} + F_{234} + F_{124} + F_{123}) \\ & + (F_{12} + F_{14} + F_{13} + F_{23} + F_{24} + F_{34}) \\ & - (F_1 + F_2 + F_3 + F_4). \end{aligned}$$

This ANOVA-type decomposition is used as a measure of dependence in the context of survival analysis [6].

All three decompositions are additive in the sense that the density is decomposed and aggregated on the probability scale. Note that there does not exist any restriction on the function F that is defined on the lattice, and that therefore the additivity condition on the probability scale is not necessary. Since the multiplicative model is the method of choice in categorical analysis [3,5], it is logical to apply a log transform to the function F. The following result suggests the equivalence of log-transformed decompositions over the monotypic, partition, and subset lattices.

2.3. Equivalence results

We first give the proof of equivalence between $\triangle^{(B)} \log F$ and $\triangle^{(M)} \log F$.

Theorem 2.1.

$$\triangle^{(B)} \log F = \triangle^{(M)} \log F.$$

Proof. Consider the coefficient of $\log F_C$, $C \subset N$ in the expansion of $\triangle^{(M)} \log F$. Observe that for $|C| \geqslant 2$, the coefficient of $\log F_C$ in the Lancaster's expansion is $(-1)^{n-|C|}$. On the other hand, by collecting coefficients, we obtain the following coefficient for $\log F_c$, where c is a singleton:

$$(-1)^{n-1}(n-1) + \sum_{j=2}^{n-1} (-1)^{n-j} \binom{n-1}{j} = (-1)^{n-1},$$

which completes the proof. □

The next theorem states that the multiplicative measures on the partition and the subset lattices are equivalent.

Theorem 2.2.

$$\triangle^{(P)} \log F = \triangle^{(B)} \log F.$$

For $n = 4$, using $\log F$ in the place of F in (5) results in the following expression:

$$\begin{aligned} \triangle^{(P)}_{1234} \log F = & \log F_{1234} - (\log F_{12|34} + \text{other terms with 2 blocks}) \\ & + 2(\log F_{12|3|4} + \text{other terms with 3 blocks}) - 6 \log F_{1|2|3|4}. \end{aligned}$$

On collecting coefficients, the right-hand side reduces to $\triangle^{(B)}_{1234} \log F$. The formal proof follows.

Proof. We require the following lemma in combinatorics.

Lemma 2.1.

$$\sum_{\pi \in \mathbf{P}(N)} (-1)^{|\pi|} |\pi|! = (-1)^n, \tag{6}$$

where $|\pi|$ is the number of blocks of π.

We are not aware of any proof of Lemma 2.1 in the literature. The proof we give is an application of involution, which is described in the context of virtual and k-species by Chen and Yeh [4].

Consider a partition $\pi = \pi_1|\pi_2|\cdots|\pi_k$. There are $|\pi|!$ linear orders of the form $\pi_{j_1}|\pi_{j_2}|\cdots|\pi_{j_k}, j_k \in \{1, \ldots, k\}$. Denote the family of linear orders of π by $\mathscr{P}(\pi)$. Assign a weight of $(-1)^{|\pi|}$ to each of the linear orders. It is clear that the left-hand side of (6) is the sum of weights in $\bigcup \mathscr{P}(\pi)$. Assuming the endowed ordering in N, we "pair off" elements in $\bigcup \mathscr{P}(\pi)$ by the following operations:

S1. Find the maximal element n in $\tau \in \bigcup \mathscr{P}(\pi)$. If it is not a singleton, then splinter n off to form a new block. That is,

$$\tau_1|\cdots|\{\tau_j \backslash n, n\}|\cdots|\tau_k \to \tau_1|\cdots|\tau_j|n|\cdots|\tau_k.$$

S2. If n is a singleton but is not the first element, then

$$\tau_1|\cdots|\tau_j|n|\cdots|\tau_k \to \tau_1|\cdots|\{\tau_j, n\}|\cdots|\tau_k.$$

S3. If n is a singleton but is the first element, then keep n in place, find the next maximal element $n-1$, and repeat the S1 and S2, and so on.

As an example, $1|234 \to 1|23|4$, $234|1 \to 23|4|1$, both by S1; $1|2|4|3 \to 1|24|3$ by S2; $4|2|13 \to 4|2|1|3$ by S3 and S1.

All elements in $\bigcup \mathscr{P}(\pi)$, except for $n|n-1|\cdots|1$, pair off to sum to zero. The weight of the only "unpaired" element or fixed point, $n|n-1|\cdots|1$ is, of course, $(-1)^n$.

To establish the equivalence of the measures on the lattices $P(N)$ and $B(N)$, we consider $\log f_C$ for a fixed $C \subset N$ and partitions of the form $C|C_1|\cdots|C_k = C|\tau$. One finds, on collecting coefficients, that $\log f_C$ has a coefficient equal to $\sum_\tau (-1)^{|\tau|}|\tau|!$ where the summation is over all possible partitions in $N \backslash C$. By Lemma 2.1, the coefficient of $\log f_C$ is $(-1)^{n-|C|}$. $\square$

3. Example and discussion

Although the equivalence decomposition over the three lattices resembles the representation in the loglinear model [3], they are distinct in their meanings. The lattice decomposition is recursively defined for every marginal distribution by virtue of Eq. (1). Consider the trivariate distribution of (X_1, X_2, X_3). The lattice decomposition of the density is jointly specified by the following set of equations:

$$\begin{aligned} \log F_{123} &= \triangle_{123}^{(B)} + \triangle_{12}^{(B)} + \triangle_{13}^{(B)} + \triangle_{23}^{(B)} \\ &\quad + \triangle_{1}^{(B)} + \triangle_{2}^{(B)} + \triangle_{3}^{(B)}, \\ \log F_{12} &= \triangle_{12}^{(B)} + \triangle_{1}^{(B)} + \triangle_{2}^{(B)}, \\ \log F_1 &= \triangle_{1}^{(B)}, \end{aligned}$$

plus the equations for $\log F_{13}$, $\log F_2$ and so on. In the above system of equations, the quantity $\triangle_{12}$ in F_{123} is equal to $\triangle_{12}$ in F_{12} for all values of (X_1, X_2). In the loglinear model, on the other hand, suppose that $\log p_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$, and that the logarithm of its marginal distribution for

(X_1, X_2) is $\log p_{ij} = u^\star + u^\star_{1(i)} + u^\star_{2(j)} + u^\star_{12(ij)}$, $u^\star_{1(i)} \neq u_{1(i)}$, $u^\star_{2(j)} \neq u_{2(j)}$, and $u^\star_{12(ij)} \neq u_{12(ij)}$. The consistency of meaning across marginal distribution, or the reproducibility property, is important in the analysis of clustered data such as those collected from car accidents, for example. In such cases, it is desirable to use a representation such that the meaning of the correlation between the degrees of injury to the driver and the passenger seated in the front is consistent across cars with different numbers of passengers. The equivalence representation therefore offers a potentially useful tool for such applications.

References

[1] M. Aigner, Combinatorial Theory, Springer, Berlin, 1979.

[2] R. Bahadur, A representation of the joint distribution of responses to n dichotomous outcomes, in: H. Solomon (Ed.), Studies in Item Analysis and Prediction, Stanford University Press, California, 1961, pp. 158–168.

[3] Y.M.M. Bishop, S.E. Fienberg, P.W. Holland, Discrete Multivariate Analysis, MIT Press, Cambridge, MA, 1975.

[4] B. Chen, Y. Yeh, Some explanation of Dobinski's formula, Stud. Appl. Math. 92 (1994) 191–199.

[5] J. Darroch, T. Speed, Additive and multiplicative models and interactions, Ann. Statist. 11 (1983) 724–738.

[6] D. Dabrowska, Weak convergence of a product integral dependence measure, Scand. J. Statist. 23 (1996) 551–590.

[7] G.M. Fitzmaurice, N.M. Laird, A.G. Rotnitzky, Regression models for discrete longitudinal responses, Statist. Sci. 8 (1993) 284–309.

[8] S.J. Haberman, The Analysis of Frequency Data, University of Chicago Press, Chicago, 1974.

[9] E.H. Ip, Y.J. Wang, Y. Yeh, Structural decompositions of multivariate density and applications in moment and cumulant, J. Multivariate Anal., to appear.

[10] H. Lancaster, The Chi-Squared Distribution, Wiley, London, 1969.

[11] K. Liang, S. Zeger, B. Qaqish, Multivariate regression analyses for categorical data (with discussion), J. R. Statist. Soc. Ser. B 54 (1992) 3–40.

[12] B. Streitberg, Lancaster interactions revisited, Ann. Statist. 18 (1990) 1878–1885.

[13] B. Streitberg, Exploring interactions in high-dimensional tables: a bootstrap alternative to log-linear models, Ann. Statist. 27 (1999) 405–413.