Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 89 (2004) 119–134
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Structural decompositions of multivariate distributions with applications in moment and cumulant

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

[a] Department of Public Health Sciences, Wake Forest University School of Medicine, Medical Center Boulevard, 3rd Floor, MRI Building, Winston-Salem, NC 27157, USA
[b] Department of Mathematical Sciences, Rutgers University, USA
[c] Institute of Mathematics, Academia Sinica, Taiwan

Received 13 October 2000

Abstract

We provide lattice decompositions for multivariate distributions. The lattice decompositions reveal the structural relationship between the Lancaster/Bahadur model and the model of Streitberg (Ann. Statist. 18 (1990) 1878). For multivariate categorical data, the decompositions allows modeling strategy for marginal inference. The theory discussed in this paper illustrates the concept of reproducibility, which was discussed in Liang et al. (J. Roy. Statist. Soc. Ser. B 54 (1992) 3). For the purpose of delineating the relationship between the various types of decompositions of distributions, we develop a theory of polytypefication, the generality of which is exploited to prove results beyond interaction.

AMS 2000 subject classifications: 62E10; 62H20

Keywords: Lattice decomposition; Lancaster model; Bahadur model; Streitberg's interaction; Cumulant; Polytypefication

*Corresponding author. Fax: 336-716-6427.
E-mail addresses: eddie.ip@marshall.usc.edu (E.H. Ip), yuwang@crab.rutgers.edu (Y.J. Wang), mayeh@math.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.

doi:10.1016/j.jmva.2003.09.001

1. Introduction

The notion of interaction among a set of variables has intrigued statisticians for many years. It has been the subject of study in interaction models [6], loglinear models [4], graphical models [33], and models of dependency [15]. Interactions describe the dependence among a set of variates. Specifically, given a distribution of multivariate variables, interactions are quantities that partition the total departure from stochastic independence. There are two approaches to interaction: the multiplicative approach that is discussed as early as [3], and later extensively studied in, among others, [4,9,10], and the additive approach discussed in [2,18]. Darroch and Speed [6] provided a general framework for representing these two types of interactions. Recently, Ip and Wang [12] discussed a lattice-based approach for reparameterizing a multiplicative loglinear model.

The traditional approach to analyze discrete data is multiplicative, as exemplified by the loglinear model [4]. However, as McCullagh [23] and others (e.g., [20]) have pointed out, the loglinear model is not designed to address questions where interest lies in the marginal distributions. Specifically, loglinear models are not reproducible [8]. That is, the parameters of the marginal distributions do not form a proper subset of the parameters of the joint distribution. For example, for the trivariate binary variable $X = (X_1, X_2, X_3)$ with cell probabilities $p_{ijk} > 0,\ i, j, k = 0, 1$, the saturated loglinear model can be written as $\log p_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)}$. The logarithm of its marginal distribution for (X_1, X_2) is $\log p_{ij} = u^{\star} + u^{\star}_{1(i)} + u^{\star}_{2(j)} + u^{\star}_{12(ij)}$, but $u^{\star}_{1(i)} \neq u_{1(i)}$, $u^{\star}_{2(j)} \neq u_{2(j)}$, and $u^{\star}_{12(ij)} \neq u_{12(ij)}$.

Because non-reproducible interactions of the loglinear model have different meanings across different marginal distributions, it is not suitable for analyzing multivariate data such as clustered categorical response of different cluster sizes. One example of responses with different cluster size is in longitudinal analysis for multiple treatment groups, where subjects within each group are observed for varying number of times. Suppose the focus is on comparing associations between responses across treatment groups. Under a non-reproducible model, parameters for different clusters cannot be directly compared. One way to solve the problem is to parameterize the joint distribution in terms of reproducible interactions, which is a motivation for this paper. For the purpose of illustration, consider multiple responses that can be grouped together to form a three-way table. Liang et al. [20] specified the joint density by the following components: three one-dimensional frequencies, three Pearson correlation coefficients, and one three-way loglinear interaction. These model components are rather different in nature: a one-dimensional frequency is non-parametric distribution, Pearson correlation, which requires scores to be assigned, is an additive interaction (see Section 4), and the three-way loglinear interaction is a multiplicative measure. In this paper we discuss three different additive decompositions of a multivariate distribution. One of the decompositions includes the Pearson correlation coefficient. All the resulting interactions of each of the three decompositions are reproducible—that is, the joint density and any of its marginal densities share the same interactions of appropriate orders. In other words, the joint density is linked to each of the marginal densities through reproducible parameters. These interactions can indeed be used to examine the compatibility among various marginal densities. Interested readers are directed to Wang [32].

The reproducible models discussed in this paper include the additive models developed by Lancaster [18] and Bahadur [2]. Recent applications of the Lancaster/Bahadur model that exploit its reproducibility property for marginal inference in data analysis can be found in [5,34]. The literature on reproducible interactions beyond two dimensions is sparse, which could be a reason why most analyses on clustered categorical responses are focused only on bivariate correlation. Consider repeated measurements (Y_1, Y_2, Y_3, Y_4) over four successive time points for treatment group 1 and five repeated measurements for treatment group 2. Associations among (Y_1, Y_2, Y_3) and (Y_2, Y_3, Y_4) for both groups may be of interest and require modeling. Using additive decomposition, a researcher can model all $2 \times (32) = 6$ two-dimensional interactions, two three-dimensional interactions, and leave the remaining higher-order additive interactions as zero functions. Unlike loglinear models, when the additive model fits well, estimates of marginal associations across treatment groups can be directly compared. Teugels and Van Horebeek [31] proposed an algebraic (additive) approach that parameterizes discrete densities in terms of moments, resulting in a decomposition appropriate for describing longitudinal data and making marginal inference. Related parameterizations are discussed in [7,30].

Streitberg [28,29] started a renewed effort to investigate properties of the Lancaster/Bahadur model. Let $N = \{1, 2, \ldots, n\}$, where n is the dimension of the distribution, and P_n be the set of all possible partitions of N. Based on a lattice theory of the P_n, Streitberg introduced a new additive model that is meant to correct a deficiency of the Lancaster/Bahadur model: for $n \geqslant 4$, the Lancaster/Bahadur interaction is shown to fail an interaction axiom, namely, if the distribution function can be factorized into a product of marginal distributions, then the highest order interaction is identically zero. Streitberg's interactions in P_n can be directly linked to cumulants.

In the first part of this paper, we present a lattice decomposition of the Lancaster/Bahadur interaction and show how their model is structurally related to Streitberg's. We develop a theory of polytypefication for this purpose. In the second part, we demonstrate how understanding the underlying structures of the two models via polytypefication facilitates proofs of results in moment and cumulant. Finally, we discuss a third model, which is based upon the Boolean algebra lattice, for density decomposition, and further establish moment and cumulant results for this model.

First, we fix some notation. Let S denote a finite set and $\prec$ be a partial order relationship defined on S. The lattice $\mathscr{L}(S)$, when exists, is formed from the ordered pair $(S, \prec)$ and has a maximal element $\hat{1}_S$, or simply $\hat{1}$ when there is no ambiguity, and a minimal element $\hat{0}_S$, or $\hat{0}$. For $\tau \in \mathscr{L}(S)$, $\Delta(\tau)$ denotes a real-valued function defined on S. The *sum function* g of Δ at σ is given by

$$g(\sigma) = \sum_{\tau \prec \sigma} \Delta(\tau). \tag{1}$$

The calculus of inversion of (1) on $\mathscr{L}(S)$ is determined by the Möbius inversion function μ. Specifically,

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

Alternatively, we can construct an operator Δ_σ for each $\sigma \in \mathscr{L}(S)$ and write (2) in shorthand as $\Delta_\sigma g = \sum \mu(\tau, \sigma) g_\tau$, sum over $\tau \prec \sigma$. We also write in shorthand $\Delta g := \Delta_{\hat{1}} g$. Existence and uniqueness results for μ are described in standard text, such as [1,27].

Eqs. (1) and (2) are dual in nature. We keep g general and may use (1), or equivalently (2), as the defining axiom for generating a class of functionals that includes interactions and moments.

2. Lattice representation of Lancaster/Bahadur measure

In this section the Lancaster/Bahadur additive interaction is shown to admit a lattice representation of form (2). Such a connection is not immediate, perhaps because it is somewhat obscured by the following symbolic definition of the interaction of Lancaster's [18]:

$$\Delta F = \prod_i (F_i^{\star} - F_i),$$

where in the expansion, $F_{i_1 \ldots i_k}$, the marginal distribution of $(X_{i_1}, \ldots, X_{i_k})$, is understood to substitute the product $F_{i_1}^{\star} \ldots F_{i_k}^{\star}$. For example, $(F_1^{\star} - F_1)(F_2^{\star} - F_2) = F_{12} - F_1 F_2$. As a further example, suppose three variables X_1, X_2, and X_3 are all binary each taking value of 0 or 1, and let $P(0,0,0) = 0.2$, $P(1,0,0) = P(0,1,0) = 0.05$, $P(0,0,1) = 0.3$, $P(1,1,0) = P(0,1,1) = P(1,0,1) = P(1,1,1) = 0.1$, where $P()$ is the cell probability, then the interaction term $\Delta F(0,0,0) = F_{123} - F_{12}F_3 - F_{23}F_1 - F_{13}F_2 + 2F_1F_2F_3 = 0.2 - (0.5)(0.4) - (0.25)(0.65) - (0.25)(0.65) + 2(0.65)(0.65)(0.4) = 0.013$. Furthermore, for the two-dimensional marginal distributions, $\Delta_{12}F(0,0) = 0.0775$, and $\Delta_{13}F(0,0) = \Delta_{23}F(0,0) = -0.01$. Here the subscript in Δ indicates the variables involved in the marginal distribution.

To see the connection between the symbolic notation and the lattice representation, we first identify a monotypic subset M_n of P_n, the set of all possible partitions of N. By this we mean if $\pi = \pi_1 | \pi_2 | \cdots | \pi_k$ is a partition of k blocks, i.e., π_j are non-empty, disjoint subsets of N, $j = 1, \ldots, k$, then

$$M_n = \{\pi \in P_n \mid \pi \text{ has at most one non-singleton block}\}.$$

A partial order $\prec$ is defined on M_n by the refinement relation between non-singleton blocks: $\pi \prec \sigma$ if and only if $A_\pi \subset A_\sigma$, where A_π denotes the non-singleton block of π. Unique infimum and supremum exist for every $\sigma, \pi \in M_n$. Thus M_n forms a lattice $\mathscr{L}(M_n)$ under $\prec$. Fig. 1 shows the structure of $\mathscr{L}(M_3)$.

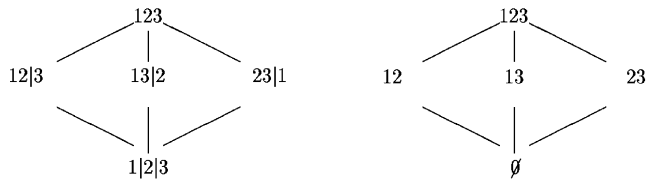

Fig. 1. Isomorphism between monotypic lattice and deatomized Boolean algebra lattice for $n = 3$.

Proposition 2.1. *The Möbius function* μ *on* $\mathscr{L}(M_n)$, *denoted by* $\mu^{(M)}$, *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.

Proof. For $\mu(\hat{0}, \hat{1}) = (-1)^{n-1}(n-1)$, we prove by induction on n. The case is trivial for $n = 2$. For $n > 2$, we use the fact that $\sum_\tau \mu(\hat{0}, \tau) = 0$. This implies

$$\mu(\hat{0}, \hat{1}) = -\sum_{\tau \in M_n^\star} \mu(\hat{0}, \tau), \tag{3}$$

where $M_n^\star = \{\tau \in M_n | \hat{0} \precneqq \tau \prec \hat{1}\}$. See Stanley [27, Section 3.10]. But for $\tau \in M_n^\star$, $\mu(\hat{0}, \tau) = \prod \mu(\hat{0}_{\tau_i}, \tau_i)$, where τ_i are the blocks of τ. For singleton τ_i, $\mu(\hat{0}_{\tau_i}, \tau_i) = 1$. Hence, $\mu(\hat{0}, \tau) = \mu(\hat{0}_A, \hat{1}_A)$, where A is the non-singleton block, $n > |A| \geqslant 2$. In (3), besides $\mu(\hat{0}, \hat{0})$, there are $\binom{n}{|A|}$ nodes whose Möbius inversion functions are $(-1)^{|A|-1}(|A|-1)$, $|A| = 2, \ldots, (n-1)$, by virtue of the induction assumption. Hence, we have

$$\mu(\hat{0}, \hat{1}) = -\left(1 + \sum_{j=2}^{n-1} (-1)^{j-1}(j-1)\binom{n}{j}\right),$$

which reduces to $(-1)^{n-1}(n-1)$.

For $\sigma \neq \hat{0}$, one can also prove by induction, but we instead make use of the observation that $\mathscr{L}(M_n)$ is isomorphic to the deatomized sublattice of the Boolean algebra 2^N of all subsets of N, where the term deatomize refers to the removing of atoms (subsets of singletons) from the lattice. This is exemplified by Fig. 1 for $n = 3$. For $\sigma \neq \hat{0}$, the Möbius inversion function, therefore, has the well-known form $\mu(\sigma, \hat{1}) = (-1)^{|\sigma|-1}$. See the section on Boolean algebra lattice. □

For an arbitrary function g defined on $\mathscr{L}(M_n)$, define a monotypic operator $\Delta^{(M)}$ such that

$$\Delta^{(M)} g = \sum_{\sigma \in M_n} \mu^{(M)}(\sigma, \hat{1}) g_\sigma, \tag{4}$$

and $\Delta_\pi^{(M)} g = \prod \Delta_{\pi_i}^{(M)} g_{\pi_i}$, $\pi \in M_n$.

The following proposition identifies Lancaster's interaction with the monotypic lattice $\mathscr{L}(M_n)$ for $n \geqslant 2$.

Proposition 2.2. *Let $n \geqslant 2$, $\pi \in M_n$ and A be the non-singleton block in π. Then*

$$\Delta_\pi^{(M)} F = \prod_{i \in A} (F_i^\star - F_i) \prod_{j \in N \setminus A} F_j,$$

and in particular, $\Delta^{(M)} F = \prod (F_i^\star - F_i)$.

Proof. Let $C \subset A$. Without loss of generality, assume $A = \{1, \ldots, |A|\}$.

$$\begin{aligned} \prod_{i \in A} (F_i^\star - F_i) &= \sum_{k=2}^{|A|} \sum_{|C|=k} (-1)^{|A|-k} F_C \prod_{j \notin C} F_j + (-1)^{|A|}(|A|-1) F_1 \ldots F_{|A|}. \\ &= \sum_\sigma \mu^{(M)}(\sigma, \hat{1}_A) F_\sigma, \end{aligned}$$

where the summation in the last expression is over σ in the monotypic sublattice generated by the set A. The proof is completed by noting that $\Delta_a^{(M)} F_a = F_a$, where a is a singleton. □

The lattice formulation leads to the following proposition.

Proposition 2.3.

$$F_N = \sum_{A \subset N} \prod_{i \in A} (F_i^\star - F_i) \prod_{j \in N \setminus A} F_j. \tag{5}$$

Proof. The Möbius inversion of (4) with $g = F$ gives $F_N = \sum \Delta_\pi^{(M)} F$. □

3. Polytypefying the Lancaster/Bahadur interaction

Identifying the Lancaster/Bahadur interaction with the monotypic lattice enables a unified treatment to study the Lancaster/Bahadur and the Streitberg's interactions. Streitberg [28] points out that the Lancaster/Bahadur interaction $\Delta^{(M)} F$ fails the interaction axiom for $n \geqslant 4$. For example, when $F_{1234} = F_{12} F_{34}$, $\Delta^{(M)} F = (F_{12} - F_1 F_2)(F_{34} - F_3 F_4)$, which in general is not identically zero. It is revealing to see how the interaction property of Lancaster/Bahadur's is limited by its lattice structure. While $\Delta^{(M)} F$ fails the "general" interaction axiom, it does satisfy a weak form of it: if $F = F_\pi$, where $F_\pi = \prod F_{\pi_i}$, $\pi \in M_n$, then $\Delta^{(M)} F = 0$. So $F_{1234} = F_{12|3|4} = F_{12} F_3 F_4$ implies $\Delta_{1234}^{(M)} F = 0$.

Now to see Streitberg's correction, define the partial ordering by refinement of $\prec$ on the full partition lattice P_n: For $\sigma, \pi \in P_n$, we write $\sigma \prec \pi$ if each block of σ is contained in a block of π. The Möbius inversion function of the partition lattice

$\mathscr{L}(P_n)$ is well-known [1]:

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

The following equation

$$\Delta^{(P)}F = \sum_{\sigma} \mu^{(P)}(\sigma, \hat{1})F_{\sigma},$$

identifies Streitberg's interaction operator $\Delta^{(P)}$. Furthermore, for $\pi \in P_n$, $\Delta_{\pi}^{(P)}F = \prod \Delta_{\pi_i}^{(P)}F_{\pi_i}$.

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

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

whilst Lancaster/Bahadur's is

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

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

It is instructive to envision $\prod(F_i^{\star} - F_i)$ as a monochromatic coloring scheme. For example, when $n = 4$, the factors in $F_1^{\star} \dots F_4^{\star}$ merge into one color to form F_{1234}, and $F_1^{\star}F_2^{\star}F_3^{\star}F_4 = F_{123}F_4$. An alternative multicoloring scheme is to consider the term $(F_1^{\star} - F_1)(F_2^{\star} - F_2) \times (F_3^{\star} - F_3)(F_4^{\star} - F_4)$ as two separate colors so that $F_1^{\star}F_2^{\star}$ merge into one color as F_{12}, while $F_3^{\star}F_4^{\star}$ merge into another as F_{34}.

Following the above heuristic, we extend the Lancaster/Bahadur measure to $\mathscr{L}(P_n)$ by defining a "polychromatic" (polytypic) operator $\Delta^{(\tilde{M})}$ using the symbolic notation: for $A \subset N$,

$$\Delta_A^{(\tilde{M})}F = \prod_{j \in A} (F_j^{\star} - F_j),$$

and for $\pi \in P_n$, $\Delta_{\pi}^{(\tilde{M})}F = \prod_i \Delta_{\pi_i}^{(\tilde{M})}F_{\pi_i}$. For example, $\Delta_{12|34}^{(\tilde{M})}F = (F_{12} - F_1F_2)(F_{34} - F_3F_4)$. A further example is

$$\underset{12|345}{\overset{(\tilde{M})}{\Delta}} F = (F_{12} - F_1F_2)[F_{345} - (F_{34}F_5 + F_{35}F_4 + F_{45}F_3) + 2F_3F_4F_5].$$

Note that if a is a singleton, $\Delta_a^{(\tilde{M})}F_a = 0$, whereas $\Delta_a^{(M)}F_a = F_a$.

We shall now prove that Streitberg's interaction is a linear combination of the polytypic Lancaster/Bahadur interactions. To this end, we define an operator $\Delta^{(\tilde{P})}F$:

$$\Delta_\pi^{(\tilde{P})} = \begin{cases} \Delta_\pi^{(P)}F & \text{if } \pi \text{ has no singleton component,} \\ 0 & \text{otherwise.} \end{cases}$$

The motivation for constructing $\Delta^{(\tilde{P})}$ is to obtain an expression for the sum function $\Delta_\sigma^{(\tilde{M})}F = \sum_{\tau \prec \sigma} \Delta_\tau^{(\tilde{P})}F$ that holds for *every* node σ on $\mathscr{L}(P_n)$. This condition enables one to make use of the Möbius inversion. The following theorem relates $\Delta^{(\tilde{M})}$ and $\Delta_\pi^{(\tilde{P})}$ and serves as a prelude to the polytypefication theorem. Proof is given in Appendix.

Theorem 3.1.

$$\Delta^{(\tilde{M})}F = \sum_{\pi \in P_n} \Delta_\pi^{(\tilde{P})}F. \tag{6}$$

Let $\sigma \in P_n$, $D(\sigma) = \{\sigma \prec \pi \mid \pi \text{ has no singleton component}\}$. Our main theorem can now be stated as follows:

Theorem 3.2 (The polytypefication theorem). *For* $\pi \in P_n$,

$$\Delta_\pi^{(\tilde{P})}F = \sum_{\sigma \prec \pi} \mu^{(P)}(\sigma, \pi)\Delta_\sigma^{(\tilde{M})}F. \tag{7}$$

In particular,

$$\Delta^{(P)}F = \sum_{\sigma \in D(\hat{0})} (-1)^{|\sigma|-1}(|\sigma| - 1)!\Delta_\sigma^{(\tilde{M})}F. \tag{8}$$

Proof. Let $\sigma \prec \pi$ and $\sigma_i = \sigma \cap \pi_i$. As an example, if $\sigma = 123|45|6|7|8$, $\pi_1 = 12345$, $\pi_2 = 678$, then $\sigma_1 = 123|45$, and $\sigma_2 = 6|7|8$.

From Theorem 3.1,

$$\Delta_{\pi_i}^{(\tilde{M})}F_{\pi_i} = \sum_{\sigma_i \prec \pi_i} \Delta_{\sigma_i}^{(\tilde{P})}F_{\sigma_i}.$$

Therefore,

$$\begin{aligned} \Delta_\pi^{(\tilde{M})}F &= \prod_{i=1}^{|\pi|} \Delta_{\pi_i}^{(\tilde{M})}F_{\pi_i} \\ &= \prod_{i=1}^{|\pi|} \sum_{\sigma_i \prec \pi_i} \Delta_{\sigma_i}^{(\tilde{P})}F_{\sigma_i} \\ &= \sum_{\sigma \prec \pi} \prod_{i=1}^{|\pi|} \Delta_{\sigma_i}^{(\tilde{P})}F_{\sigma_i} \\ &= \sum_{\sigma \prec \pi} \Delta_\sigma^{(\tilde{P})}F_\sigma. \end{aligned}$$

Eq. (7) follows from Möbius inversion and (8) from the definition of $\Delta^{(\tilde{P})}$. □

As an example, when $n = 4$,

$$\Delta^{(P)}F = \prod_{j=1}^{4} (F_j^{\star} - F_j) - \left[\prod_{j=1,2} (F_j^{\star} - F_j) \prod_{j=3,4} (F_j^{\star} - F_j) + \prod_{j=1,3} (F_j^{\star} - F_j) \prod_{j=2,4} (F_j^{\star} - F_j) + \prod_{j=1,4} (F_j^{\star} - F_j) \prod_{j=2,3} (F_j^{\star} - F_j) \right].$$

The polytypefication theorem reveals an interesting relationship between the structures of the monotypic and the full partition lattices. Their Möbius inversion functions can be explicitly related by the expression

$$\mu^{(P)}(\tau, \hat{1}) = \sum \mu^{(P)}(\sigma, \hat{1}) \mu^{(\tilde{M})}(\tau, \sigma),$$

where the sum is over $\sigma \in D(\tau)$. This connection is independent of the sum function. In other words, the same connection holds for interactions for density functions, distribution functions and other potential functions. The general nature of the polytypefication theorem can therefore be exploited—treat moment and cumulant as sum functions defined on lattice structures to prove new results.

4. An application of polytypefication to moment and cumulant

Moments, central moments, and cumulants have been extensively studied for their roles in characterizing distribution functions [17]. In particular, the relation between cumulants and the partition lattice is investigated in [24–28]. McCullagh [22] contains some general tensor methods for cumulants.

To fix notation, we denote the expected value of X_j by α_j, central moment $E \prod_i (X_i - \alpha_i)$ by ξ_N, moment $E \prod_i X_i$ by α_N, and the corresponding cumulant by κ_N. Define the moment product $\alpha_\pi = \prod_i E(X_{\pi_i})$ so for instance, $\alpha_{13|2} = E(X_1 X_3) E(X_2)$ [28]. From the definitions of cumulant [24] and central moment, we have

$$\kappa_N = \Delta^{(P)} \alpha,$$

and

$$\xi_N = \Delta^{(\tilde{M})} \alpha.$$

The analogy of κ_N and ξ_N to $\Delta^{(P)}F$ and $\Delta^{(\tilde{M})}F$ is immediate: the sum function is moment instead of distribution function. The polytypefication theory described in Section 3 applies without much modification and leads to the following corollary:

Corollary 4.1.

$$\kappa_N = \Delta^{(P)}\alpha = \sum_{\sigma \in D(\hat{0})} (-1)^{|\sigma|-1}(|\sigma|-1)!\Delta_\sigma^{(\tilde{M})}\alpha$$
$$= \sum_{\sigma \in D(\hat{0})} (-1)^{|\sigma|-1}(|\sigma|-1)!\xi_\sigma.$$

In other words, cumulant is a linear combination of polytypic central moments, $\Delta_\sigma^{(\tilde{M})}\alpha$.

Lehmann [19] proves that when $n=2$, $\xi_N = \int(F_{12} - F_1F_2)$. In the following proposition, we prove that the result generalizes to $n>2$: central moment is a signed integral of the Lancaster/Bahadur multivariate interaction. Our result thus connects moment and interaction, both of which are used extensively in the literature to parameterize multivariate density [6,31]. As we shall see, this integration theorem readily extends to the full partition lattice $\mathscr{L}(P_n)$ via the polytypefication theorem for moments. But first we state a generalization of Lehmann's integration theorem.

Proposition 4.1.

$$\xi_N = (-1)^n \int \Delta^{(\tilde{M})} F(\boldsymbol{u})\, d\mathbf{u},$$

if the integral on the right-hand side exists.

Proof. Trivial when $n=1$. For $n \geqslant 2$, let $X^\bigstar = (X_1^\bigstar, \ldots, X_n^\bigstar)$ denote a random vector with cdf $F(x)$ and $F_A(x_A)$ denote the marginal cdf of $X_A^\bigstar$. Furthermore, suppose X_i is an independent and identical copy of $X_i^\bigstar$, and X_i is stochastically independent of X_j for every $i \neq j$. Then

$$\xi_N = E\prod_{i=1}^n (X_i^\bigstar - \alpha_i)$$
$$= E\prod_{i=1}^n (X_i^\bigstar - X_i)$$
$$= E\int \prod_{i=1}^n [I(x_i, u_i) - I(x_i^\bigstar, u_i)]\, d\mathbf{u},$$

where $I(x,u) = 1$ when $x \leqslant u$ and 0 otherwise. Under the assumption that integration and expectation freely exchange, we have

$$\xi_N = \int E\left[\prod_{i=1}^n (I(x_i, u_i) - I(x_i^\bigstar, u_i))\right] d\mathbf{u}$$
$$= (-1)^n \int \Bigg\{ \sum_{A \subset N, |A| \geqslant 2} (-1)^{n-|A|} P(X_A^\bigstar \leqslant u_A) \prod_{j \in N \setminus A} P(X_j \leqslant u_j)$$

$$+(-1)^{n-1}\sum_{i} P(X_i^{\star} \leqslant u_i)\prod_{j\neq i} P(X_j \leqslant u_j) + (-1)^n \prod_{i=1}^{n} P(X_i \leqslant u_i)\Bigg\} d\mathbf{u}$$
$$=(-1)^n \int \Bigg\{ \sum_{A\subset N, |A|\geqslant 2} (-1)^{n-|A|} F_A(u_A) \prod_{j\in N\setminus A} F_j(u_j) + (-1)^{n-1}(n-1)\prod_{i=1}^{n} F_i(u_i)\Bigg\} d\mathbf{u}$$
$$=(-1)^n \int \Delta^{(\tilde{M})} F(\mathbf{u})\, d\mathbf{u}. \qquad \square$$

The theorem can be immediately extended, using standard product operator notation, to product central moment.

Corollary 4.2. *For* $\pi \in P_n$,

$$\xi_\pi = \Delta_\pi^{(\tilde{M})}\alpha = (-1)^n \int \Delta_\pi^{(\tilde{M})} F(\mathbf{u})\, d\mathbf{u}, \tag{9}$$

if the integral on the right-hand side exists.

Finally, an application of polytypefication to Eq. (9) yields the following relationship between cumulant and interaction.

Proposition 4.2.

$$\kappa_N = (-1)^n \int \Delta^{(P)} F(\mathbf{u})\, d\mathbf{u},$$

if the integral on the right-hand side exists.

Proof.

$$(-1)^n \int \Delta^{(P)} F(\boldsymbol{u})\, d\mathbf{u} = \int (-1)^n \sum_{\sigma\in D(\hat{0})} \mu^{(P)}(\sigma,\hat{1}) \Delta_\sigma^{(\tilde{M})} F(\mathbf{u})\, d\mathbf{u} \quad \text{(from (8))}$$
$$= \sum_{\sigma\in D(\hat{0})} \mu^{(P)}(\sigma,\hat{1}) \int (-1)^n \Delta_\sigma^{(\tilde{M})} F(\mathbf{u})\, d\mathbf{u}$$
$$= \sum_{\sigma\in D(\hat{0})} \mu^{(P)}(\sigma,\hat{1}) \Delta_\sigma^{(\tilde{M})}\alpha \quad \text{(from Corollary 4.2)}$$
$$= \Delta^{(P)}\alpha. \quad \text{(from Corollary 4.1)} \qquad \square$$

5. Extension to include Boolean algebra lattice

Understanding the underlying structure of interactions in the monotypic and full partition lattices allows the expansion of the scope of lattice decomposition to include a new structure, the Boolean algebra lattice (see [14]). Let B_n be the Boolean algebra of N and define an order relation $\prec$ on the power set 2^N so that $C \prec D$ if and only if $C \subset D$. The power set thus forms a lattice $\mathscr{L}(B_n)$ with joint $C \vee D = C \cup D$ and meet $C \wedge D = C \cap D$ for any $C, D \in 2^N$. Hence, $\hat{1} := N$, and $\hat{0} := \emptyset$.

The Möbius inversion function of $\mathscr{L}(B_n)$ is

$$\mu^{(B)}(D, C) = (-1)^{|C|-|D|},$$

where $|D|$ denotes the cardinality of D [1].

Define an additive interaction measure $\Delta^{(B)}F$ on $\mathscr{L}(B_n)$ by (1). Hence for $n = 2$, $F_{12} = \Delta_{12}^{(B)} + \Delta_1^{(B)} + \Delta_2^{(B)} + \Delta_\emptyset^{(B)}$. We take $F_\emptyset = 1$ and from Möbius inversion $\Delta^{(B)}F = F_{12} - F_1 - F_2 + 1$. More generally,

$$\Delta^{(B)}F = \sum_{A \subset N} (-1)^{n-|A|} F_A.$$

The interaction function $(-1)^n \Delta^{(B)}F$ is the multivariate survival function $\bar{F}$ of an n-vector variable. See Joe [15, p. 10]. The operator $\Delta^{(B)}$ is idempotent and satisfies $\Delta^{(B)}F = (-1)^n \bar{F}$, and $\Delta^{(B)}\bar{F} = (-1)^n F$. It is easy to verify that $\Delta_A^{(B)}F_A = \prod_{j \in A} (F_j^\bigstar - 1)$. This lemma follows from the symbolic representation:

Lemma 5.1.

$$\Delta^{(\tilde{M})}F = \sum_{A \subset N} \Delta_A^{(B)} F_A \prod_{j \notin A} (1 - F_j),$$

$$\Delta^{(B)}F = \sum_{A \subset N} \Delta_A^{(\tilde{M})} F_A (-1)^{n-|A|} \prod_{j \notin A} (1 - F_j).$$

Proof.

$$\begin{aligned} \Delta^{(\tilde{M})}F &= \prod_{j=1}^{n} (F_j^\bigstar - F_j) \\ &= \prod_{j=1}^{n} [(F_j^\bigstar - 1) + (1 - F_j)] \\ &= \sum_{A \subset N} \Delta_A^{(B)} F_A \prod_{\notin A} (1 - F_j). \end{aligned}$$

The proof for $\Delta^{(B)}F$ follows from the expansion in $\Delta^{(B)}F = \prod (F_j^\bigstar - 1) = \prod [(F_j^\bigstar - F_j) - (1 - F_j)]$. $\square$

The lemma leads to yet another integration result: for multivariate positive variates, moments are signed integrals of survival functions. The result is a multivariate generalization of the equality that if $X_1 \geqslant 0$, then $EX_1 = \int \bar{F}(x_1)\, dx_1$.

Proposition 5.1. *Suppose* $X_i \geqslant 0$ *for all* $i = 1, \ldots, n$. *If* $\int \Delta_A^{(\tilde{M})} F_A \, d\mathbf{u}_A$ *exist for* $A \subset N$, *and* EX_j *exist for* $j = 1, \ldots, n$, *then*

$$\alpha_N = (-1)^n \int \Delta^{(B)} F(\mathbf{u}) \, d\mathbf{u}$$

Proof. From the definition of moment, one has

$$\alpha_N = E \prod (X_i - \alpha_i + \alpha_i) = \sum_{A \subset N} \xi_A \prod_{j \notin A} \alpha_j.$$

Hence

$$\begin{aligned} \alpha_N &= \sum_{A \subset N} (-1)^{|A|} \int_A \Delta_A^{(\tilde{M})} F_A(\mathbf{u}_A) \, d\mathbf{u}_A \prod_{j \notin A} \alpha_j \\ &= (-1)^n \int_A \sum_{A \subset N} \left\{ (-1)^{n-|A|} \Delta_A^{(\tilde{M})} F_A(\mathbf{u}_A) \, d\mathbf{u}_A \int_{N \setminus A} \prod_{j \notin A} (1 - F_j) \, d\mathbf{u}_{N \setminus A} \right\}, \end{aligned}$$

from which the theorem follows by virtue of Lemma 5.1. □

6. Discussion

An important mathematical technique used in this paper is the Möbius inversion—a familiar and basic tool in lattice theory. By using the Möbius inversion, we have avoided the use of flat [31] and tensor [21] for representing cumulants. Teugels [30] pointed out that matrix and kronecker products "failed in finding a relative easy tensor formulation" for cumulants. In this paper, the integration theorem (Proposition 4.1), which is based on the polytypefication theorem, presents a relatively straightforward formulation.

Our discussion so far has been restricted to multivariate discrete variables, but actually the lattice theory developed in this paper is applicable to continuous variables. We choose to emphasize discrete variable because issues such as reproducibility that one encounters in multivariate discrete analysis do not require as much attention in continuous variable [8]. The lattice theory might well facilitate multivariate analysis of continuous variables in areas such as the study of local dependence function [11,16].

The restriction of our discussion to additive models can also be relaxed. The lattice and polytypefication theories apply to general sum functions. Therefore, by taking logarithm of the joint and marginal densities, one can directly construct multiplicative models. Some work has already been started in that direction [13,14].

Appendix. Proof of the polytypefication theorem

Before proceeding to prove the polytypefication result in Section 3, we need some notations in combinatorics and a lemma.

Let $\mathbb{S}(A)$ denote the set of permutations of A, and we write $\mathbb{S}(N)$ as $\mathbb{S}_n$. For any $\tau \in \mathbb{S}_n$, τ can be decomposed into disjoint cyclic permutations τ_i, $i = 1, \ldots, |\tau|$. Denote the parity of τ_i by $\omega(\tau_i)$, i.e., $\omega(\tau_i) = (-1)^{|\tau_i|-1}$, where $|\tau_i|$ is the cardinality of τ_i. Let $\omega(\tau) = \prod \omega(\tau_i)$. For a basic treatise on the subject of permutation, see [21].

Now let $C(A)$ denote the set of all circular permutations of A. Suppose $\pi = \pi_1|\pi_2|\cdots|\pi_k$ is a partition of N. So π can be associated to a collection of permutations

$$S(\pi) = \{\tau = \tau_1 \ldots \tau_i \ldots \tau_k \mid \tau_i \in C(\pi_i)\},$$

where $|S(\pi)| = \prod_i (|\pi_i| - 1)!$. For $\mathscr{A} \subset P_n$, let

$$S(\mathscr{A}) = \{\tau \in \mathbb{S}_n \mid \tau \in S(\pi) \text{ for some } \pi \in \mathscr{A}\}.$$

The next lemma follows from the above definitions.

Lemma A.1.

$$\sum_{\tau \in S(\mathscr{A})} \omega(\tau) = \sum_{\pi \in \mathscr{A}} \prod_{i=1}^{|\pi|} (-1)^{|\pi_i|-1}(|\pi_i| - 1)!.$$

The proof of Theorem 3.1 proceeds as follows.

When $n = 1$, both sides are zero by definition. Assume $n \geqslant 2$. The right-hand side of (6) can be rewritten as

$$\begin{aligned}\sum_{\pi \in D(\hat{0})} \Delta_\pi^{(P)} F &= \sum_{\pi \in D(\hat{0})} \sum_{\sigma \prec \pi} \mu^{(P)}(\sigma, \pi) F_\sigma \\ &= \sum_{\sigma \in P_n} \sum_{\pi \in D(\sigma)} \mu^{(P)}(\sigma, \pi) F_\sigma.\end{aligned}$$

For a given σ, let $\pi \in D(\sigma)$ and $\sigma_i = \sigma \cap \pi_i$.

From Lemma A.1, we have

$$\sum_{\sigma \in P_n} \sum_{\pi \in D(\sigma)} \prod_{i=1}^{|\pi|} \mu^{(P)}(\sigma_i, \pi_i) F_\sigma = \sum_{\sigma \in P_n} \sum_{\tau \in S(D(\sigma))} \omega(\tau) F_\sigma.$$

In the example $\sigma = 123|45|6|7|8$, $\pi = 12345|678$, $\sigma_1 = 123|45$, and $\sigma_2 = 6|7|8$, the corresponding $\omega(\tau) = (-1)^{2-1}(-1)^{3-1} = -1$.

Denote $\sum_{\tau \in S(D(\sigma))} \omega(\tau)$ by $\Omega_\sigma(k,m)$, where k, m are respectively the number of non-singleton and singleton blocks in σ. We claim

$$\Omega_\sigma(k,m) = \begin{cases} (-1)^{m-1}(m-1), m \geqslant 2 & \text{if } k = 0, \\ (-1)^m, m \geqslant 0 & \text{if } k = 1, \\ 0, m \geqslant 0 & \text{if } k > 1. \end{cases}$$

The proof is accomplished by induction. The case for $k + m = 1$ is trivial: $\Omega_\sigma(1,0) = 1$. Assume $k + m \geqslant 2$.

Case 1: $k = 0$, $\Omega_\sigma(0,2) = -1$. When $m \geqslant 2$, we have $\Omega_\sigma(0,m) = -(m-1)\Omega_\sigma(1,m-2) = (-1)^{m-1}(m-1)$.

Case 2: $k = 1$, $\Omega_\sigma(1,1) = -1$. Assume $m \geqslant 2$, $\Omega_\sigma(1,m) = \Omega_\sigma(0,m) - m\Omega_\sigma(1,m-1) = (-1)^m$.

Case 3: $k = 2$, $\Omega_\sigma(2,0) = 0$. For $k \geqslant 3$,

$$\Omega_\sigma(k,m) = \Omega_\sigma(k-1,m) - (k-1)\Omega_\sigma(k-1,m) - m\Omega_\sigma(k,m-1).$$

It follows then, by induction assumption, that $\Omega_\sigma(k,m) = 0$. The claim is thus proved, and it can be seen that $\Omega_\sigma(k,m)$ is exactly the coefficient of F_σ in $\Delta^{(\tilde{M})}F$.

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.

[3] M.S. Bartlett, Contingency table interactions, J. Roy. Statist. Soc. Suppl. 2 (1935) 248–252.

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

[5] A. Carlier, P.M. Kronenberg, Decompositions and biplots in three-way correspondence analysis, Psychometrika 61 (1996) 353–373.

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

[7] A. Ekholm, P. Smith, J. McDonald, Marginal regression analysis of a multivariate binary response, Biometrika 82 (1995) 847–854.

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

[9] L.A. Goodman, The multivariate analysis of qualitative data: interactions among multiple classifications, J. Amer. Statist. Assoc. 65 (1970) 225–256.

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

[11] P. Holland, Y. Wang, Dependence functions for continuous bivariate densities, Comm. Statist. Theory Methods 16 (1987) 863–976.

[12] E.H. Ip, Y.J. Wang, A strategy for designing telescoping models for analyzing multiway contingency tables using mixed parameters, Sociol. Methods Res. 31 (2003) 291–324.

[13] E.H. Ip, Y.J. Wang, Y. Yeh, Reproducible lattice models for multivariate interactions, Technical Report IOM99-01, Marshall School of Business, University of Southern California, California, 1999.

[14] E.H. Ip, Y.J. Wang, Y.N. Yeh, Some invariance results concerning multiplicative decompositions of a multivariate density, J. Multivariate Anal. 84 (2003) 403–409 doi:10.1016/S0047-259X(02)00051-9.

[15] H. Joe, Multivariate Models and Dependence Concepts, Chapman & Hall, London, 1997.

[16] M.C. Jones, Constant local dependence, J. Multivariate Anal. 16 (1998) 148–155 doi:10.1006/jmva.1997.1714.
[17] M. Kendall, A. Stuart, The Advanced Theory of Statistics, Vol. 1, Hafner, New York, 1963.
[18] H. Lancaster, The Chi-Squared Distribution, Wiley, London, 1969.
[19] E. Lehmann, Some concepts of dependency, Ann. Math. Statist. 37 (1966) 1137–1153.
[20] K. Liang, S. Zeger, B. Qaqish, Multivariate regression analyses for categorical data (with discussion), J. Roy Statist. Soc. Ser. B 54 (1992) 3–40.
[21] P. MacMahon, Combinatory Analysis; Cambridge University Press, Cambridge, MA, reprinted by Chelsea, New York, 1960.
[22] P. McCullagh, Tensor Methods in Statistics, Chapman & Hall, New York, 1987.
[23] P. McCullagh, Models for discrete multivariate responses, Bull. I.S.I. Paris 53 (1989) 407–418.
[24] T. Speed, Cumulants and partition lattices, Austral. J. Statist. 25 (1983) 378–388.
[25] T. Speed, Cumulants and partition lattices ii: generalized k-statistics, J. Austral. Math. Soc. Ser. A 40 (1986a) 34–53.
[26] T. Speed, Cumulants and partition lattices iii: multiply indexed arrays, J. Austral. Math. Soc. Ser. A 40 (1986b) 161–182.
[27] R.P. Stanley, Enumerative Combinatorics, Vol. 1. Wadsworth, Monterey, CA, 1986.
[28] B. Streitberg, Lancaster interactions revisited, Ann. Statist. 18 (1990) 1878–1885.
[29] B. Streitberg, Exploring interactions in high-dimensional tables: A bootstrap alternative to log-linear models, Ann. Statist. 27 (1999) 405–413.
[30] J.L. Teugels, Some representations of multivariate Bernoulli and binomial distributions, J. Multivariate Anal. 32 (1990) 256–268.
[31] J.L. Teugels, J. Van Horebeek, Algebraic descriptions of nominal multivariate discrete data, J. Multivariate Anal. 67 (1998) 203–226 doi:10.1006/jmva.1998.1764.
[32] Y.J. Wang, Compatibility among marginal densities, Biometrika, in press.
[33] J. Whittaker, Graphical Models in Applied Mathematical Multivariate Statistics, Wiley, Chichester, 1990.
[34] A. Whittemore, Logistic regression of family data from case–control studies, Biometrika 82 (1995) 57–67.