The Annals of Statistics
1990, Vol. 18, No. 4, 1878–1885

LANCASTER INTERACTIONS REVISITED

BY BERND STREITBERG

University of Hamburg

Additive interactions of n-dimensional random vectors X, as defined by Lancaster, do not necessarily vanish for $n \geq 4$ if X consists of two mutually independent subvectors. This defect is corrected and an explicit formula is derived which coincides with Lancaster's definition for $n < 4$. The new definition leads also to a corrected Bahadur expansion and has certain connections to cumulants. The main technical tool is a characterization theorem for the Moebius function on arbitrary finite lattices.

1. Introduction: A defect in Lancaster's definition of interaction measures. The concept of additive interaction measures was introduced by Lancaster (1969), although a special case already appear in Bahadur's (1961) representation of a multidimensional probability distribution concentrated on $\{0, 1\}^n$. An (additive) interaction measure ΔF is a signed measure determined by a given distribution F on $\mathbb{R}^n$ which vanishes identically whenever F in the nontrivial product of two of its (multivariate) marginal distributions. Besides having an independent value as a structural concept, interaction measures have many applications in statistics, e.g.:

1. contingency table analysis, cf. Zentgraf (1975), Toewe, Bock and Kundt (1985);
2. statistical physics [cf. Falkenhagen and Ebeling (1963)];
3. a new axiomatization of cumulants, cf. Section 4 of this paper. This axiomatization suggests several ways for the robustification of cumulants.

Lancaster (1969) introduces an interaction measure as a certain signed measure ΔF, a function of the joint distribution $F = F_{1,2,\dots,n}$ of a random vector $X = (X_1, \dots, X_n)$ that is supposed to be identically zero whenever the random vector is decomposable into two mutually independent subvectors. For $n = 2$, he finds

$$\Delta F(x_1, x_2) = F_{12}(x_1, x_2) - F_1(x_1) F_2(x_2),$$

or, in a symbolic notation,

$$\Delta F = (F_1^* - F_1)(F_2^* - F_2),$$

where after expansion a product like $F_i^* F_j^* \cdots F_k^*$ is understood to denote the corresponding joint distribution function $F_{i,j,\dots,k}$, cf. also Darroch and Speed [(1983), page 736]. The interaction measure ΔF for general n then is

Received April 1989; revised November 1989.
AMS 1980 *subject classifications*. 62E10, 62E30.
Key words and phrases. Additive interactions, Bahadur expansions, cumulants, Moebius function, partition lattice, contingency tables.

given by

$$\Delta F = (F_1^* - F_1)(F_2^* - F_2) \cdots (F_n^* - F_n).$$

While this definition has the intended property for $n = 3$ where

$$\Delta F = F_{123} - F_1 F_{23} - F_2 F_{13} - F_3 F_{12} + 2F_1 F_2 F_3,$$

it does not work as intended for $n \geq 4$. Assuming that $X = (X_1, X_2, X_3, X_4)$ is decomposable into the two mutually independent subvectors (X_1, X_2) and (X_3, X_4), one finds after expansion and simplification

$$\Delta F = (F_{12} - F_1 F_2)(F_{34} - F_3 F_4),$$

which is, in general, unequal to the zero function.

We will, in the following, define a corrected version of Lancaster's interaction measure. The symbolic notation will not be used, the main technical tool being a characterization of the Moebius function on an arbitrary finite lattice. The new definition of interaction measures will lead immediately to a corrected Bahadur representation of probability measures. In a precise sense, Lancaster's original definition corresponds to moments about the means, while the new definition introduces measures which correspond to cumulants. In his work, Lancaster gave not only the measure theoretical definition [Lancaster (1969), Chapter XII, 2] but also an alternative approach to interaction measures via L^2 expansions. These expansions are essentially moment generating functions. While not discussed in the following paper, an analogous L^2 theory can easily be developed for the new interaction measures. The expansions then will correspond to multivariate cumulant generating functions.

2. Partitions and interactions. A *partition* π of $\mathbf{n} = \{1, 2, \ldots, n\}$ is a set of nonempty, pairwise disjoint subsets $\pi_i \subseteq \mathbf{n}$, called the *blocks* of the partition, the union of which is equal to $\mathbf{n}$. We write, as a shorthand notation, instead of, e.g., $\{\{1, 3\}, \{2\}, \{4\}\}$, simply $13|2|4$. The set of all partitions π of $\mathbf{n}$ is denoted by $\mathbf{P(n)}$. The lattice structure on $\mathbf{P(n)}$ is most easily described by viewing a partition as an equivalence relation on $\mathbf{n}$. Let $\mathbf{R}$ be the set of reflexive and symmetric relations on $\mathbf{n}$, i.e., subsets $R \subseteq \mathbf{n} \times \mathbf{n}$ with $(i, i) \in R$ for all $i \in \mathbf{n}$ and $(i, j) \in R$ whenever $(j, i) \in R$. The equivalence relations are precisely the closed elements $R = R^*$ in $\mathbf{R}$, where R^* is the transitive closure of R. We identify an equivalence relation with the partition given by its equivalence classes. For partitions τ, π we define $\tau \leq \pi$ if $\tau \subseteq \pi$ as relations and the lattice operations by $\tau \wedge \pi := \tau \cap \pi$ and $\tau \vee \pi := (\tau \cup \pi)^*$. The finest partition is $0 := 1|2| \cdots |n$ and the coarsest one is $1 := 1, 2, \ldots, n$. The lattice of partitions appears naturally in many classical and, since the rebirth of invariant theory, modern parts of statistics and combinatorics [cf. Carney (1968), Rota (1964), Tracy and Gupta (1973), Speed (1983), Stanley (1986)].

The following construction of *partition operators* is exemplified by the simple operator $J_{13|2}$, which transforms a three-dimensional distribution function $F_{123} = F(x_1, x_2, x_3)$ into the product $F_{13|2} = F_{13|2}(x_1, x_2, x_3) =$

$F(x_1, \infty, x_3)F(\infty, x_2, \infty)$. Given an n-dimensional distribution function $F = F_{1,2,\dots,n}$ of a random vector $(X_1, \dots, X_n)$ and a partition $\pi = \pi_1|\pi_2|\cdots|\pi_k \in P(n)$, let F_π be the product of the marginal distribution functions of the k subvectors $(X_i\colon i \in \boldsymbol{\pi}_j)$, $j = 1, 2, \dots, k$. The map $F \to F_\pi$ is a (nonlinear) operator J_π on $\mathbb{M}_1$, the set of probability measures on $(\mathbb{R}^n, \mathscr{B}_n)$, and the operators J_π have a simple composition rule which is easy to prove:

PROPOSITION 1. *For* $\tau, \pi \in \mathbf{P(n)}$: $J_\tau \circ J_\pi = J_{\tau \wedge \pi}$.

The intended characteristic property of Lancaster interactions can be defined as follows:

DEFINITION 1. A measure $F \in \mathbb{M}_1$ is called *decomposable* if $\pi \in \mathbf{P(n)}$ with $\pi < 1$ exists such that $J_\pi F = F$. An (additive) *interaction operator* is a linear combination

$$\Delta = \sum_\pi a_\pi J_\pi,$$

with $\Delta F = 0$ for *all* decomposable measures F. A measure $F \in \mathbb{M}_1$ is called τ-*decomposable* for $\tau \in \mathbf{P(n)}$ if $\pi \in \mathbf{P(n)}$ with $\pi < \tau$ exists such that $J_\pi F = F$. A τ-interaction operator Δ_τ is a linear combination

$$\Delta_\tau = \sum_\pi a(\pi, \tau) J_\pi$$

such that $\Delta_\tau F = 0$ for *all* τ-decomposable F. We set $\Delta_0 = J_0$.

Clearly, $\Delta = \Delta_1$, i.e., $a_\pi = a(\pi, 1)$. Note that F is τ-decomposable if $F = F_\tau$ and furthermore, at least one of the subvectors defined by the blocks τ_j of τ: $(X_i\colon i \in \tau_j)$ is itself decomposable. An interaction operator Δ_τ is a map from $\mathbb{M}_1$ to $\mathbb{M}$, the (Banach) space of all finite signed measures on $(\mathbb{R}^n, \mathscr{B}_n)$, i.e., the *interaction measure* ΔF is a finite signed measure. For $n = 2$, Definition 1 immediately gives, up to a multiplicative factor, $\Delta = \Delta_{12} = J_{12} - J_{1|2}$, i.e., Lancaster's definition for the case $n = 2$. With a bit more algebra, one also finds directly from the definition that ΔF as defined here coincides with Lancaster's result for $n = 3$.

3. A corrected version of Lancaster interactions and Bahadur expansions. In the derivation of the general result, *Iverson's convention*, cf. Graham, Knuth and Patashnik (1989), will be used: given a proposition $\mathscr{P}$, let $\{\mathscr{P}\}$ denote its truth value, i.e., $\{\mathscr{P}\}$ is equal to 1 if $\mathscr{P}$ is true and equal to 0 if $\mathscr{P}$ is false.

For all τ, $\sigma \in \mathbf{P(n)}$ and all $\gamma < \tau$, the sum of $a(\pi, \tau)$ over all π with $\pi \wedge \gamma = \sigma$ vanishes. Assume namely that $J_\gamma F = F$ for an F with $J_\tau F = F$ and $\gamma < \tau$, i.e., F is τ-decomposable. Since $\Delta_\tau F$ has to vanish for all such F, an equality of operators has to hold: $\sum a(\pi, \gamma) J_{\pi \wedge \gamma} = 0$. The statement follows by collecting coefficients.

We show that $a(\cdot, \cdot)$ is, up to a constant, uniquely defined as the Moebius function on the partition lattice. In order to state this in general terms, let S be an arbitrary finite lattice and define the *incidence algebra* $I(S)$ as the set of all mappings g: $S \times S \to \mathbb{R}$ with $g(x, y) \neq 0$ only if $x \leq y$. We introduce on $I(S)$ the operations of

1. scalar multiplication, for $c \in \mathbb{R}$: $(cg)(x, y) = cg(x, y)$;
2. addition, $(g + h)(x, y) = g(x, y) + h(x, y)$;
3. convolution, $(gh)(x, y) = \sum_z g(x, z)h(z, y)$.

It is easy to show, that $I(S)$ is closed under these operations, cf. Aigner (1975), and is, indeed, an associative $\mathbb{R}$-linear algebra (which can be realized as a subalgebra of the algebra of upper triangular matrices). Special elements of $I(S)$ are the *identity* δ with $\delta(x, y) = \delta_{xy}$, the *zeta function* ζ with $\zeta(x, y) = \{x \leq y\}$ and the Moebius function μ as the inverse of the zeta function, i.e., $\mu\zeta = \zeta\mu = \delta$. The following basic lemma characterizes the Moebius function:

LEMMA 1. *Let S be a finite lattice. In the incidence algebra $I(S)$ there exists exactly one function a: $S \times S \to \mathbb{R}$ with the following properties*:

(i) $a(x, x) = 1$ *for all* $x \in S$,
(ii) $\sum_x \{x \wedge t = s\} a(x, y) = 0$ *for all* $y, s \in S$ *and all* $t < y$,

and $a = \mu$, *the Moebius function in* $I(S)$.

PROOF. (i) Uniqueness. Let $a, b \in I(S)$ satisfy (i) and (ii). Consider $e = a - b$. We want to show $e = 0$. For each $y \in S$, one has $e(y, y) = a(y, y) - b(y, y) = 0$ by (1). We proceed by an induction-type argument. Let $S(y)$ be the set of all $z < y$ for which $e(z, y) = 0$ has not yet been shown. If $S(y)$ is nonempty, it contains at least one maximal element, say t. Set $s = t$ in (ii), then (summations over all $x \in S$):

$$\begin{aligned} 0 = \sum \{x \wedge t = t\} e(x, y) &= \sum \{t \leq x \leq y\} e(x, y) \\ = e(t, y) + \sum \{t < x \leq y\} e(x, y) &= e(t, y), \end{aligned}$$

because the sum in the last line vanishes by definition of t. Therefore $e(t, y) = 0$, hence, by induction, $e = 0$.

(ii) Existence. We show that μ satisfies (i) and (ii). By definition,

$$\sum_y \{x \leq y \leq z\} \mu(x, y) \zeta(y, z) = \delta(x, z),$$

hence $\mu(x, x) = \delta(x, x) = 1$. Choose $y \in S$ and $t < y$. Then (summations over $x \in S$):

$$\begin{aligned} \sum \{x \wedge t = t\} \mu(x, y) &= \sum \{t \leq x \leq y\} \mu(x, y) \\ &= \sum \{x \leq y\} \zeta(t, x) \mu(x, y) = \delta(t, y) = 0, \end{aligned}$$

i.e., (ii) for $s = t < y$. Now let $S(t, y)$ be the set of all $s \in S$ with $s < t < y$ for which (ii) has not yet been shown for μ. If $S(t, y)$ is nonempty, it has at least one maximal element s. One finds (summations over $x \in S$):

$$\begin{aligned}\sum \{x \wedge t = s\}\mu(x, y) &= \sum \{x \wedge t \geq s\}\mu(x, y) - \sum \{x \wedge t > s\}\mu(x, y) \\ &= \sum \{x \wedge t \geq s\}\mu(x, y) \quad \text{(by definition of } s\text{)} \\ &= \sum \{x \geq s\}\mu(x, y) \\ &= \sum \zeta(s, x)\mu(x, y) = \delta(s, y) = 0.\end{aligned}$$

The statement follows by induction. □

A special case of the existence part of Lemma 1 [that μ satisfies (ii) for $y = 1$, $s = t < y$] can already be found in Weisner (1935), the author who first generalized the original method by Moebius (1832) to àrbitrary partial orders. Here, however, both the general case and the uniqueness are crucial.

PROPOSITION 2. *In the definition of the interaction operator Δ_τ, the coefficient $a(\tau, \tau)$ can be chosen equal to* 1. *Then, uniquely,*

$$\Delta_\tau = \sum \mu(\pi, \tau) J_\pi,$$

and, in particular,

$$\Delta = \sum (-1)^{|\pi| - 1}(|\pi| - 1)! J_\pi,$$

where $|\pi|$ denotes the number of blocks of a partition π.

PROOF. Immediately from Lemma 1 and the well known explicit form for the Moebius function on the partition lattice [e.g., Speed (1983)]. □

EXAMPLE 1. For $n = 4$, ΔF is given by

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

Note that Δ_τ, for $\tau < 1$, is a product of interaction operators applied to the marginal distributions defined by the blocks of τ, e.g.,

$$\Delta_{12|34} F = (F_{12} - F_1 F_2)(F_{34} - F_3 F_4).$$

PROPOSITION 3. *An arbitrary distribution function F has the Bahadur-type expansion*:

$$F = \sum_\tau \Delta_\tau(F).$$

PROOF. Definition of μ. □

Because J_π operates on densities in the same way as on measures, Propositions 2 and 3 can immediately be restated in a density version whenever F

affords a density with respect to counting or Lebesgue measure (say in order to define interaction in multidimensional contingency tables).

EXAMPLE 2. In this example, we use the simplified and not completely rigorous notation of tensor algebra for the representation of multidimensional arrays. Let $f = (f_{ijkl})$ be a four-dimensional contingency table with positive one-dimensional marginal distributions f_i, f_j, f_k, f_l and write, e.g., $r_{ij}r_{kl}$ for the table corresponding to $(\Delta_{12|34} f_{ijkl})/f_i f_j f_k f_l$. Note that $\Delta_{1|2|3|4} f$ is equal to the product $f_i f_j f_k f_l$ of the one-dimensional marginal distributions. The Bahadur-type expansion of f is as follows:

$$f_{ijkl} = f_i f_j f_k f_l(1 + r_{ij} + r_{ik} + r_{il} + r_{jk} + r_{jl} + r_{kl} + r_{ij}r_{kl} + r_{ik}r_{jl} + r_{il}r_{jk} + r_{ijk} + r_{ijl} + r_{ikl} + r_{jkl} + r_{ijkl}).$$

In the original Bahadur expansion, the product terms (e.g., $r_{ij}r_{kl}$) are absent.

Statistical comments: A common application of Bahadur-type expansions is the approximate representation of an empirical contingency table by low order interaction terms, cf. Zentgraf (1975). While the estimation theory of truncated Bahadur representations is still largely undeveloped, it is possible that the appearance of further multiplicative components might produce superior approximations. An alternative approach for contingency table analysis is based on the bootstrap and will be discussed in a forthcoming paper.

4. Interactions and cumulants. Interaction measures have an interesting connection to cumulants. We give an axiomatic approach to these quantities. Let $\mathbb{M}_1^n \subseteq \mathbb{M}_1$ be the set of probability measures F on $\mathbb{R}^n$ for which $E[|X_1 X_2 \cdots X_n|] = \int |x_1 x_2 \cdots x_n| \, df$ exists. The cumulant $\kappa(X_1, X_2, \dots, X_n) = \kappa(X)$ can be interpreted as a real-valued function on $\mathbb{M}_1^n$ which is characterized by five properties (expressed, with a slight abuse of notation, in terms of random variables):

PROPERTY 1. Symmetry. $\kappa(X) = \kappa(X^\sigma)$ for all permutations σ of $\mathbf{n}$, where $(X^\sigma) := (X_{\sigma(1)}, X_{\sigma(2)}, \dots, X_{\sigma(n)})$.

PROPERTY 2. Multilinearity. $\kappa(aX_1, X_2, \dots, X_n) = a\kappa(X_1, X_2, \dots, X_n)$ and $\kappa(X_1 + Y_1, X_2, \dots, X_n) = \kappa(X_1, X_2, \dots, X_n) + \kappa(Y_1, X_2, \dots, X_n)$.

PROPERTY 3. Moment property. $\kappa(X) = \kappa(Y)$ if X and Y have identical mixed moments up to order n.

From Properties 2 and 3, it follows that κ is a linear combination of the *moment products* $M_\pi(X) = \int x_1 x_2 \cdots x_n \, dJ_\pi F$, where for instance $M_{13|2}(X) = E[X_1 X_3]E[X_2]$.

PROPERTY 4. Normalization. The coefficient of κ in $M_1 = E[X_1 \cdots X_n]$ is equal to 1.

PROPERTY 5. Interaction property. $\kappa(X) = 0$ if X is decomposable.

PROPOSITION 4. *κ is uniquely given by Properties 1–5 as*

$$\kappa(X) = \sum (-1)^{|\pi|-1}(|\pi| - 1)! M_\pi(X) = \int x_1 x_2 \cdots x_n \, d\Delta F,$$

where F is the distribution function of X.

PROOF. Immediate from Lemma 1 and Proposition 2. □

The fact that κ can be written as an integral with respect to a signed measure ΔF suggests many ways of defining κ-type parameters in a robust fashion, e.g., via replacing the coordinate functions x_i by bounded functions $h_i(x_i)$. Even without a proof of Proposition 4 (e.g., Lemma 1), low-order cumulants can be computed very simply from Properties 1–5 and also the well-known properties of cumulants (translation invariance, addition theorem, cumulants of linear functions, cumulants in the multivariate normal distribution, etc.) are easy consequences of Properties 1–5. The axiomatic approach has, therefore, certain pedagogic virtues. If one applies the theory of generating functions by Doubilet, Stanley and Rota (1972), also the role of log $E[e^{i(t,x)}]$ becomes transparent. In this sense, the approach to cumulants from interaction measures is dual to Speed's (1983) discussion of these quantities.

5. Multiplicative and additive interaction models. Multiplicative interactions in contingency tables are usually represented within the framework of loglinear ANOVA-type models. A *term* of such a model corresponds to a *subset* $T \subseteq \mathbf{n}$, where $\mathbf{n}$ indexes the dimensions of the table. The set of terms has a natural lattice structure as the Boolean algebra $2^{\mathbf{n}}$ of all subsets of $\mathbf{n}$. A *hierarchical model* is given by a subset $\mathscr{H} \subseteq 2^{\mathbf{n}}$ with the property that with $T \in \mathscr{H}$ also $S \in \mathscr{H}$ for all $S \subseteq T$, i.e., $\mathscr{H}$ is an *order ideal* of $2^{\mathbf{n}}$ [cf. Aigner (1975) or Stanley (1989) for the basic concepts of order theory]. Alternatively, $\mathscr{H}$ can be specified by its *Sperner family*, i.e., the maximal terms in $\mathscr{H}$. The set of hierarchical models is the ideal lattice $\mathscr{I}(2^{\mathbf{n}})$ of $2^{\mathbf{n}}$, i.e., the free distributive lattice on $\mathbf{n}$. This structural characterization can, for instance, be used to *count* the number of hierarchical models (to our knowledge, this is an unsolved problem for $n > 12$).

Additive interactions, on the other hand, do not correspond to *subsets*, but to *partitions* of $\mathbf{n}$. The set of terms, therefore, is given by the partition lattice $\mathbf{P}(\mathbf{n})$. Again, a hierarchical model $\mathscr{H}$ is an order ideal of the term lattice, i.e., with $\pi \in \mathscr{H}$ also $\tau \in \mathscr{H}$ for all $\tau \leq \pi$. The set of hierarchical models is the ideal lattice $\mathscr{I}(\mathbf{P}(\mathbf{n}))$, which is again distributive. In our opinion, the distinction between subsets and partitions is the main structural difference between multiplicative and additive interaction models. While statisticians are well trained to operate intuitively within the free distributive lattice of ANOVA models, the corresponding intuition for the ideals of the partition lattice still remains to be developed.

REFERENCES

AIGNER, M. (1975). *Kombinatorik I. Grundlagen und Zaehltheorie*. Springer, Berlin.

BAHADUR, R. R. (1961). A representation of the joint distribution of responses to n dichotomous items. In *Studies in Item Analysis and Prediction* (H. Solomon, ed.) 158–176. Stanford Univ. Press.

CARNEY, E. J. (1968). Relationship of generalized polykays to unrestricted sums for balanced complete finite populations. *Ann. Math. Statist.* **39** 643–656.

DARROCH, J. N. and SPEED, T. P. (1983). Additive and multiplicative models and interactions. *Ann. Statist.* **11** 724–738.

DOUBILET, P., STANLEY, R. and ROTA, G. C. (1972). On the foundations of combinatorial theory VI. The idea of generating functions. *Proc. Sixth Berkeley Symp. Math. Statist. Probab.* **2** 267–318. Univ. California Press.

FALKENHAGEN, H. and EBELING, W. (1963). Zur kinetischen Theorie schwach ionisierter Plasmen im Magnetfeld. *Ann. Physik* (7) 347–359.

GRAHAM, R. L., KNUTH, D. E. and PATASHNIK, O. (1989). *Concrete Mathematics. A Foundation for Computer Science*. Addison-Wesley, Reading.

LANCASTER, H. O. (1969). *The Chi-Squared Distribution*. Wiley, London.

MOEBIUS, A. F. (1832). Ueber eine besondere Art der Umkehrung der Reihen. *J. Reine Angew. Math.* **9** 105–123.

ROTA, G. C. (1964). On the foundations of combinatorial theory I. Theory of Moebius functions. *Z. Wahrsch. Verw. Gebiete* **2** 340–368.

SPEED, T. P. (1983). Cumulants and partition lattices. *Austral. J. Statist.* **25** 378–388.

STANLEY, R. P. (1986). *Enumerative Combinatorics* **1**. Wadsworth, Monterey, Calif.

TOEWE, J., BOCK, J. and KUNDT, G. (1985). Interactions in contingency table analysis. *Biometrical J.* **27** 1–7.

TRACY, D. S. and GUPTA, B. C. (1973). Multiple products of polykays using ordered partitions. *Ann. Statist.* **1** 913–923.

WEISNER, L. (1935). Abstract theory of inversion of finite series. *Trans. Amer. Math. Soc.* **38** 474–484.

ZENTGRAF, R. (1975). A note on Lancaster's definition of higher order interactions. *Biometrika* **62** 375–378.

ALT-MOABIT 90
D–1000 BERLIN 21
WEST GERMANY