Discrete Mathematics 69 (1988) 61–77
North-Holland

EXISTENCE OF EXTENSIONS AND PRODUCT EXTENSIONS FOR DISCRETE PROBABILITY DISTRIBUTIONS

F.M. MALVESTUTO
STUDI DOC – E.N.E.A., 00198 Roma, Italy

Received 2 January 1986
Revised 3 December 1986

Three or more probability distributions may be pairwise compatible but not collectively compatible, in the sense that they admit no common extensions. However, pairwise compatibility proves to be a necessary and sufficient condition for collective compatibility when the underlying system of distribution schemes is "acyclic". If this is the case, then (and only then) do the distributions admit a product extension, whose expression can be computed by a simple algorithm.

1. Introduction

The question of the compatibility of a given set of discrete probability distributions is fundamental to Probability Theory and important for many problems in Statistics and in Information Science. As an immediate extension of well-known measure-theoretical results, one can prove the existence of common extensions of discrete distributions if their sample spaces are assumed to be "almost independent" [6]. Under this hypothesis, the problem has infinite solutions. However, if the extensions are required to have a "multiplicative form", then the solution is unique. When the hypothesis of independence is released, an answer for the question of compatibility is not yet known.

In this paper, we give a sufficient condition: if the distributions in question are pairwise compatible and the system of their schemes is "acyclic", then they are collectively compatible. Indeed, a stronger result is proven: acyclic scheme systems are uniquely determined as the the ones for which any family of pairwise compatible distributions is collectively compatible, too. From another point of view, acyclic systems can be characterized by the existence of *product extensions*. An immediate consequence of such a property is the possibility of constructing extensions, whose functional expressions are generated using a polynomial algorithm, known as "reduction procedure".

Section 2 and Section 3 contain the mathematical background. In Section 4 the definition of maximum-entropy extension is introduced and the computational advantages of acyclic systems of distribution schemes are shown. Section 5 deals with "product forms", whose notion is formally defined. There we shall prove

0012-365X/88/$3.50

that a product extension is nothing but a maximum-entropy extension defined over an acyclic system of distribution schemes. Section 6 contains the proof that the acyclicity of the system of the distribution schemes is a necessary and sufficient condition for the existence of a common extension of pairwise compatible distributions.

2. Distributions and extensions

Let X be a finite set of discrete variables, called *attributes*, which have associated finite sets of values. An *X-tuple* (or a *tuple*, if X is understood) is defined by a combination of possible choices of values for each attribute in X. If Y is a subset of X and x is an X-tuple, then $x(Y)$ denotes its *Y-component* obtained by discarding from x the values of the attributes not in Y.

A *discrete probability distribution* is a tern $\langle X, \Omega, p(x)\rangle$, where:

X is a finite set of attributes,

Ω is a finite set of X-tuples,

$p(x)$ is a normalized function that associates a nonnegative number with each tuple in Ω.

The sets X and Ω will be referred to respectively as the *scheme* and the *space* of the distribution. Henceforth, it is understood that Ω is the Cartesian product of the value-sets associated with the attributes in X. The subset R of Ω, defined as

$$R = \{x: p(x) > 0\},$$

will be referred to as the *characteristic relation* of the distribution.

In the following a discrete probability distribution will be specified by assigning its distribution function. Given a distribution $p(x)$ over X, we can construct a "marginal" distribution for every nonempty subset Y of X, simply taking the *restriction* of $p(x)$ to Y:

$$p(y) = \sum p(x),$$

the summation being extended over the X-tuples x with $x(Y) = y$.

Given a system of probability distributions $p_1(x_1), \ldots, p_s(x_s)$, we say that they are

pairwise compatible if the restrictions of $p_i(x_i)$ and $p_j(x_j)$ to the set of the common attributes coincide, for all i and j;

collectively compatible if there exists a common extension, that is, a distribution over $X = \bigcup_i X_i$, that has $p_1(x_1), \ldots, p_s(x_s)$ as its marginals.

The following example demonstrates that pairwise compatibility is not sufficient for collective compatibility.

Example 1. Let A, B and C be binary attributes. The second-order distributions

AB	p_1	AC	p_2	BC	p_3
00	0.5	00	0	00	0.5
01	0	01	0.5	01	0
10	0	10	0.5	10	0
11	0.5	11	0	11	0.5

turn out to be pairwise compatible, but not collectively compatible. In fact, the existence of a common extension, $p(abc)$, would give rise to the following contradiction:

$$\begin{aligned} 0.5 = p_2(01) &= p(001) + p(011) \\ &\leq [p(001) + p(101)] + [p(011) + p(010)] \\ &= p_3(01) + p_1(01) = 0. \end{aligned}$$

Testing the compatibility of a distribution system is equivalent to testing the feasibility of a linear-programming problem, known as "multi-index transportation problem" (see Section 6), for which the existence of necessary and sufficient conditions for feasibility is still an open problem. However a number of necessary conditions, in addition to the obvious conditions of pairwise compatibility, are known [9]. Among these, we want to mention the following set of inequalities, known as the *Schell conditions*:

$$p(x) \leq \min\{p_1(x_1), \ldots, p_s(x_s)\},$$

where x_i is the X_i-component of x. From the Schell conditions it follows that the characteristic relation of any extension of $p_1(x_1), \ldots, p_s(x_s)$ is a nonempty (proper or improper) subset of the *join*, R^*, of their characteristic relations, defined as

$$R^* = \{x \in \Omega: x_1 \in R_1, \ldots, x_s \in R_s\},$$

where R_i is the characteristic relation of $p_i(x_i)$. So, the nonemptiness of R^* is a necessary condition for $p_1(x_1), \ldots, p_s(x_s)$ to be collectively compatible. This explains why the distributions of Example 1 are incompatible. Indeed, the characteristic relations of collectively compatible distributions have to answer the requirement of so-called "collective consistency" [1], which is a stronger condition than the join's nonemptiness. Now we introduce the notion of collective consistency as well as the notions of pairwise consistency and independence, which are fundamental to relational data theory [1].

Given a system of schemes $S = \{X_1, \ldots, X_s\}$, for each X_i arbitrarily choose a relation R_i. We say that the relations $R_1, \ldots, R_s$ are

collectively consistent if there exists a relation over $X = \bigcup_i X_i$ (called an

"extension" of $R_1, \ldots, R_s$) such that its projection onto X_i coincides with R_i, for all i;
pairwise consistent if the projections of R_i and R_j onto their common attributes are the same, for all i and j;
independent if the cardinality of their join is equal to the product of their cardinalities.

It is easily seen that the pairwise/collective compatibility of given distributions implies the pairwise/collective consistency of their characteristic relations. It should be noticed that if the relations $R_1, \ldots, R_s$ are collectively consistent, then their join R^* is the extension with the largest number of tuples. So, in order to test the collective consistency of given relations, we have to compute their join and compare each relation with the homologous projection of the join. If all of them match, then and only then are they collectively consistent.
The following example demonstrates that the pairwise compatibility of given distributions and the collective consistency of their characteristic relations are not sufficient to assure their collective compatibility.

Example 2. Let A, B and C be second-order attributes. The bivariate distributions

AB	p_1	AC	p_2	BC	p_3
00	0.1	00	0.3	00	0.4
01	0.4	01	0.2	01	0.1
10	0.4	10	0.2	10	0.1
11	0.1	11	0.3	11	0.4

are pairwise compatible and such that the join of their characteristic relations is nonempty. Nevertheless, they are incompatible. In fact, the existence of an extension, $p(abc)$, would give rise to the following contradiction:

$$\begin{aligned} 0.4 = p_1(01) &= p(010) + p(011) \\ &\leqslant [p(010) + p(110)] + [p(001) + p(011)] \\ &= p_3(10) + p_2(01) = 0.3. \end{aligned}$$

In Section 6 we shall trace those cases where pairwise compatibility is sufficient to assume collective compatibility.

3. Acyclic scheme systems

Let $S = \{X_1, \ldots, X_s\}$ be a system of schemes. The attributes are usually put in two classes; *common* attributes, being those involved in more than one scheme; and *unique* attributes, appearing in exactly one scheme.
A scheme is *extreme* if it contains one or more unique attributes.
Consider the following algorithm, sometimes called *reduction procedure* [1, 3].

Reduction procedure

Apply the following two operations to the scheme system $S = \{X_1, \ldots, X_s\}$ repeatedly until neither can be further applied.

(a) (DELETION) Delete all the unique attributes of an extreme scheme.

(b) (REDUCTION) Delete a redundant scheme.

By "reduced system" we mean the result, S', of the reduction procedure. The redundant schemes deleted by reduction will be referred to as the *reducing factors* of S. The system S is called *acyclic* if the reduced system S' is an empty set. If this is the case, the order where the schemes in S are one by one deleted by reduction is called a *perfect reduction ordering* (p.r.o.).

Example 3. The scheme system $S = \{ABC, ABD, ACE, BCF\}$ is acyclic since it reduces to an empty set. An effective p.r.o. is: BCF, ACE, ABD, ABC.

Scheme system	Step 1	Step 2	Step 3	Reduced system
ABC	ABC	ABC	ABC	–
ABD	ABD	ABD	–	–
ACE	ACE	–	–	–
BCF	–	–	–	–

As the reduction procedure is a polynomial-time algorithm [1], testing acyclicity is an easy task (a linear algorithm can be found in [8]).

Basic properties of acyclicity:

1. *Running intersection property* [1]

A scheme system S is acyclic if and only if there is an ordering $X_1, \ldots, X_s$ of its schemes such that for each $i > 1$ there exists $j < i$ such that

$$(X_1 \cup \cdots \cup X_{i-1}) \cap X_i = X_j \cap X_i.$$

Such a permutation will be referred to as a *running intersection ordering* (r.i.o.).

The intersections such as $X_j \cap X_i$, which represent the structural links among the schemes in S, are nothing but the reducing factors of S. Moreover, if $X_1, \ldots, X_s$ is a r.i.o., then $X_s, \ldots, X_1$ is a p.r.o.

Finally, it should be noticed that in the search for an optimal permutation the choice of the first element $t(1)$ is immaterial, since for each X_i there is an ordering of S with $X_{t(1)} = X_i$, that enjoys the running intersection property.

2. *Relational consistency property* [1]

A scheme system $S = \{X_1, \ldots, X_s\}$ is acyclic if and only if every system of pairwise consistent relations over S is also collectively consistent.

Additional properties of acyclic systems can be found in [1].

4. Maximum-entropy extensions

Let $\{p_1(x_1), \ldots, p_s(x_s)\}$ be a system of distributions over $S = \{X_1, \ldots, X_s\}$ and $X = \bigcup_i X_i$. If the distributions are collectively compatible, consider the set of all possible extensions $p(x)$.

It is well-known that the two following definitions of the extension $p^*(x)$ of $p_1(x_1), \ldots, p_s(x_s)$ are equivalent [4]:

(a) $p^*(x)$ is the extension, *uniquely* determined as the one with the largest Shannon "entropy"

$$H[p(x)] = -\sum p(x) \log p(x).$$

(b) $p^*(x)$ is the extension, uniquely determined as the one factoring in the form

$$p^*(x) = a_1(x_1) \cdots a_s(x_s),$$

where the implicit functions a_i's are determined to satisfy the marginal constraints:

$$p^*(x_i) = p_i(x_i) \quad (i = 1, \ldots, s).$$

Maximum-entropy extensions can be computed by an iterative procedure, called Iterative Proportional Fitting Procedure (IPFP) [2, 4].

The following theorem stresses the connection that there exists between entropy maximization and relational join.

Theorem 1. *The characteristic relation of the maximum-entropy extension of a system of (collectively compatible) distributions is given by the join of their characteristic relations.*

Proof. As the maximum-entropy extension can be written as

$$p^*(x) = a_1(x_1) \cdots a_s(x_s),$$

the characteristic relation of $p^*(x)$ is given by the join of the characteristic relations of $a_1(x_1), \ldots, a_s(x_s)$. Now, consider the marginal constraint $p_i(x_i) = p^*(x_i)$. We have

$$p_i(x_i) = a_i(x_i) b_i(y_i),$$

where

$$b_i(y_i)=\sum a_1(x_1)\cdots a_{i-1}(x_{i-1})a_{i+1}(x_{i+1})\cdots a_s(x_s),$$

the summation being extended over all the values $x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_s$ consistent with x_i. So, $p_i(x_i)$ vanishes if $a_i(x_i)$ vanishes. This implies that the characteristic relation R_i of $p_i(x_i)$ is a subset of the characteristic relation of $a_i(x_i)$ and, therefore, that the join of $R_1,\ldots,R_s$ is a subset of the characteristic relation of $p^*(x)$.

On the other hand, as stated above, the Shell conditions imply that the characteristic relation of any extension of $p_1(x_1),\ldots,p_s(x_s)$ is a subset of the join of $R_1,\ldots,R_s$. Hence the characteristic relation of $p^*(x)$ must coincide with the join of $R_1,\ldots,R_s$. $\square$

The maximum-entropy extension takes a very simple expression in the case that $X_1,\ldots,X_s$ form a partition of $X=\bigcup_i X_i$.

Theorem 2. *Let $X_1,\ldots,X_s$ be pairwise disjoint schemes and $X=\bigcup_i X_i$. The maximum-entropy extension $p^*(x)$ of an arbitrary system of distributions $p_1(x_1),\ldots,p_s(x_s)$ over $S=\{X_1,\ldots,X_s\}$ can be written simply*

$$p^*(x)=p_1(x_1)\cdots p_s(x_s). \tag{1}$$

Proof. Let $p(x)$ be an arbitrary extension of $p_1(x_1),\ldots,p_1(x_s)$. In information theory the so-called "mutual information" [7], defined as

$$\sum_i H[p_i(x_i)]-H[p(x)],$$

is known to be a nonnegative quantity. Therefore, we have

$$H[p(x)]\leq\sum_i H[p_i(x_i)]=H[p^*(x)]. \quad \square$$

Extensions of form (1) are called *multiplicative* [6]. The following two theorems characterize multiplicative extensions and correspond to well-known results in the measure theory (see Theorem 1 and Theorem 2 in [6]).

Theorem 3. *A system $\{p_1(x_1),\ldots,p_s(x_s)\}$ of distributions admits a multiplicative extension if and only if their characteristic relations are independent.*

Proof. (if) The functional expression

$$p(x)=p_1(x_1)\cdots p_s(x_s)$$

is normalized to the unity and, therefore, defines a distribution. Moreover, it satisfies the marginal constraints.

(only if) Let $p(x)$ be a multiplicative extension. As $p(x)>0$ if and only if

$p_i(x_i) > 0$ for all i, the characteristic relation of $p(x)$ has the same cardinality as the Cartesian product of the characteristic relations of $p_1(x_1), \ldots, p_s(x_s)$. $\square$

Theorem 4. *Let $S = \{X_1, \ldots, X_s\}$ be a system of distribution schemes and $X = \bigcup_i X_i$. Each system of distributions over S admits a multiplicative extension if and only if the schemes in S are pairwise disjoint.*

Proof. It is a consequence of Theorem 3 by the light of the fact that, the pairwise disjointness of the schemes in S is a necessary and sufficient condition for the independence of each system of relations over S. $\square$

We have seen a case where the computation of the maximum-entropy extension is easy, it being enough to take the product of the component distributions.

We are now interested in tracing the general case where the functional expression of the maximum-entropy extension can be determined a priori, that is, without explicitly knowing the component distributions. Such a property is desirable for we avoid resorting to the IPFP when computing maximum-entropy extensions. We shall prove (see Theorem 7) that this is the case when the system of distribution schemes given is acyclic.

Theorem 5. *Given a scheme system $S = \{X_1, \ldots, X_s\}$, let Z_i be the set of unique attributes of X_i and $Y_i = X_i - Z_i$ be the set of its common attributes. Let $X = \bigcup_i X_i$ and $Y = \bigcup_i Y_i$. If $\{p_1(x_1), \ldots, p_s(x_s)\}$ is any system of collectively compatible distributions over S, the maximum-entropy extension $p^*(x)$ can be written as*

$$p^*(x) = \left[\prod_i p_i(x_i)/p_i(y_i)\right] p^*(y), \tag{2}$$

where $p^*(y)$ is the maximum-entropy extension of $p_1(y_1), \ldots, p_s(y_s)$, being $p_i(y_i)$ the restriction of $p_i(x_i)$ to Y_i.

Proof. Let us take the implicit form of the maximum-entropy extension

$$p^*(x) = a_1(x_1) \cdots a_s(x_s).$$

Enforcing the marginal constraint $p_i(x_i) = p^*(x_i)$, we obtain

$$p_i(x_i) = a_i(x_i) b_i(y_i),$$

where

$$b_i(y_i) = \sum a_1(x_1) \cdots a_{i-1}(x_{i-1}) a_{i+1}(x_{i+1}) \cdots a_s(x_s),$$

the summation being extended over all the values $x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_s$ that turn out to be collectively consistent with x_i. Summing $p_i(x_i)$ over all Z_i-values, one has

$$p_i(y_i) = b_i(y_i)c_i(y_i),$$

where

$$c_i(y_i) = \sum_{z_i} a_i(x_i).$$

Combining the expressions of $p_i(x_i)$ and $p_i(y_i)$, one finds

$$a_i(x_i) = [p_i(x_i)/p_i(y_i)]c_i(y_i).$$

Hence

$$p^*(x) = \left[\prod_i p_i(x_i)/p_i(y_i)\right]c(y),$$

where $c(y) = \prod_i c_i(y_i)$.

It remains to be proven that $c(y)$ coincides with the maximum-entropy extension $p^*(y)$ of $p_1(y_1), \ldots, p_s(y_s)$. On account of the fact that by definition $c(y)$ has a factorized form, it suffices to check only that $c(y)$ is a common extension of $p_1(y_1), \ldots, p_s(y_s)$, that is, for all i,

$$p_i(y_i) = \sum c(y),$$

the summation being extended over all y with Y_i-component equal to y_i. But, this follows from the fact that $p_i(y_i)$ must coincide with the restriction to Y_i of

$$p^*(x) = \left[\prod_i \frac{p_i(x_i)}{p_i(y_i)}\right]c(y). \quad \square$$

Note that (1) is a special case of (2).

The following theorem is an extension of Theorem 5.

Theorem 6. *Given a cyclic scheme system $S = \{X_1, \ldots, X_s\}$, let $\{V_h\}$ be the set of the reducing factors of S, and $S' = \{W_k\}$ be the reduced system. If $\{p_1(x_1), \ldots, p_s(x_s)\}$ is any system of collectively compatible distributions over S, the maximum-entropy extension $p^*(x)$ can be written as follows*

$$p^*(x) = \frac{\prod_i p_i(x_i)}{\prod_h p_h(v_h)} \frac{p^*(w)}{\prod_k p_k(w_k)}, \tag{3}$$

where $p^*(w)$ is the maximum-entropy extension of the distribution system $\{p_k(w_k)\}$.

Proof. Let $S_1 = \{X_{i_1}^{[1]}: i_1 = 1, \ldots, s\}$ be the system S of distribution schemes given. Given $S_m = \{X_{i_m}^{[m]}: i_m \in I_m\}$ for $m \geq 1$, the system $S_{m+1} = \{X_{i_{m+1}}^{[m+1]}: i_{m+1} \in I_{m+1}\}$

is taken by applying to S_m the two above operations of deletion and reduction. That is, if $Y_{i_m}^{[m]}$ is the subset of $X_{i_m}^{[m]}$ formed by its common attributes, then S_{m+1} is obtained by deleting the redundant elements of the set $\{Y_{i_m}^{[m]}\}$, so that $I_{m+1} \subset I_m$. If we denote by $\{Y_{j_m}^{[m]}\}$ the set of such redundant elements, one has

$$\{Y_{i_m}^{[m]}\} = \{Y_{j_m}^{[m]}\} \cup \{X_{i_{m+1}}^{[m+1]}\}.$$

Let $S_n = \{X_{i_n}^{[n]}\}$ be the last nonempty reduction. If the system S is cyclic, then each attribute of $X^{[n]}$ appears in two or more of its component schemes. Then, apply Theorem 5 repeatedly $n-1$ times. Taking into account that $\{p_{i_m}(y_{i_m}^{[m]})\}$ and $\{p_{i_{m+1}}(x_{i_{m+1}}^{[m+1]})\}$ have the same maximum-entropy extension $p^*(x^{[m+1]})$, we find

$$\begin{cases} p^*(x^{[1]}) = \left[\prod_{i_1} p_{i_1}(x_{i_1}^{[1]})/p_{i_1}(y_{i_1}^{[1]})\right] \cdot p^*(x^{[2]}), \\ p^*(x^{[2]}) = \left[\prod_{i_2} p_{i_2}(x_{i_2}^{[2]})/p_{i_2}(y_{i_2}^{[2]})\right] \cdot p^*(x^{[3]}), \\ \quad\vdots \qquad \vdots \\ p^*(x^{[n-1]}) = \left[\prod_{i_{n-1}} p_{i_{n-1}}(x_{i_{n-1}}^{[n-1]})/p_{i_{n-1}}(y_{i_{n-1}}^{[n-1]})\right] \cdot p^*(x^{[n]}). \end{cases}$$

Combining these results, we have

$$p^*(x) = \prod_i p_i(x_i) \left\{ \prod_{m=1,\ldots,n-2} \frac{\prod_{i_{m+1}} p_{i_{m+1}}(x_{i_{m+1}}^{[m+1]})}{\prod_{i_m} p_{i_m}(y_{i_m}^{[m]})} \right\} \frac{p^*(x^{[n]})}{\prod_{i_{n-1}} p_{i_{n-1}}(y_{i_{n-1}}^{[n-1]})}.$$

As $\{X_{i_{m+1}}^{[m+1]}\}$ is a subset of $\{Y_{i_m}^{[m]}\}$, then each fraction

$$\prod_{i_{m+1}} p_{i_{m+1}}(x_{i_{m+1}}^{[m+1]}) \Big/ \prod_{i_m} p_{i_m}(y_{i_m}^{[m]}),$$

reduces to

$$1 \Big/ \prod_{j_m} p_{j_m}(y_{j_m}^{[m]}).$$

Hence

$$p^*(x) = \frac{\prod_i p_i(x_i)}{\prod_{m=1,\ldots,n-2} \prod_{j_m} p_{j_m}(y_{j_m}^{[m]}) \prod_{i_{n-1}} p_{i_{n-1}}(y_{i_{n-1}}^{[n-1]})} \frac{p^*(x^{[n]})}{}.$$

If $\{V_h\}$ is the set of the interaction factors of the scheme system S, we have

$$\{V_h\} = \bigcup_{m=1,\dots,n-1} \{Y_{j_m}^{[m]}\},$$

and finally

$$p^*(x) = \frac{\prod_i p_i(x_i)}{\prod_h p_h(v_h)} \frac{p^*(x^{[n]})}{\prod_{i_n} p_{i_n}(x_{i_n}^{[n]})}, \tag{4}$$

identical to (3) after setting $X^{[n]} = W$ and $\{X_{i_n}^{[n]}\} = \{W_k\}$. □

In virtue of this theorem, in all the cases where the underlying scheme system is reducible ($S \neq S'$), we can reduce the computation cost of the maximum-entropy extension by applying the IPFP to the distributions $p_k(w_k)$ rather than to the original distributions $p_i(x_i)$. Moreover, in the case that the scheme system is acyclic, the following theorem shows that we need not resort to the IPFP since the maximum-entropy extension has a closed-form expression.

Theorem 7. *Given an acyclic scheme system $S = \{X_1, \dots, X_s\}$, let $\{V_h\}$ be the set of the reducing factors of S. If $\{p_1(x_1), \dots, p_s(x_s)\}$ is any system of collectively compatible distributions over S, the maximum-entropy extension $p^*(X)$ can be written as follows*

$$p^*(x) = \prod_i p_i(x_i) \Big/ \prod_h p_h(v_h). \tag{5}$$

Proof. The formula (4) holds also if S is acyclic. Moreover, $S_n = \{X_{i_n}^{[n]}\}$ is a partition of $X^{[n]}$ and, therefore, by Theorem 4

$$p^*(w^{[n]}) = \prod_{i_n} p_{i_n}(w_{i_n}^{[n]}).$$

Then, the formula (4) reduces to (5). □

5. Product extensions

In the previous section, we were able to give a closed-form expression to the maximum-entropy extension of any acyclic system of collectively compatible distributions. In this section, we introduce a class of functional expressions, we call *product forms*, by means of which it is easy to construct extensions of given distributions. We shall prove that the existence of a product form is a characteristic property of acyclic systems of distribution schemes. Given a scheme system $S = \{X_1, \ldots, X_s\}$, consider a functional expression such as

$$f(X) = \prod_i p_i(X_i) \Big/ \prod_h p_h(V_h),$$

where $f(X)$, $p_i(X_i)$ and $p_h(V_h)$ are distribution symbols and each V_h is a suitable (nonempty) subset of a certain X_i. We say that $f(X)$ satisfies the "unity sum property" if the following procedure, which we call *simplification procedure,* terminates with success.

Simplification procedure

Apply the following two operations to the functional expression $f(X)$ repeatedly until neither can be further applied.

(a) (DELETION) Delete all the unique attributes of an extreme scheme. If X_i reduces to an empty set, then delete the distribution symbol $p_i(X_i)$.

(b) (SIMPLIFICATION) Delete $p_i(X_i)$ and $p_h(V_h)$ if $X_i = V_h$.

If the algorithm reduces $f(X)$ to nothing, we say that it terminates *with success.*

If $f(X)$ is a functional expression that satisfies the unity sum property, then it does denote a proper distribution over X, when given $\{p_1(x_1), \ldots, p_s(x_s)\}$.

Example 4. The functional expression

$$f(ABCDE) = \frac{p_1(ABC)p_2(AD)p_3(BE)}{p_1(AB)}$$

satisfies the unity sum property.

The functional expression

$$f(X) = \prod_j p_j(X_j) \Big/ \prod_h p_h(V_h)$$

is said to satisfy the *marginal constraint* $f(X_i) = p_i(X_i)$ if the following procedure, which we call *selective simplification procedure,* terminates with success.

Selective simplification procedure

Apply the following two operations to the functional expression $f(X)$ repeatedly until neither can be further applied.

(a) (SELECTIVE DELETION) Delete all the unique attributes of an extreme scheme X_j $(j \neq i)$. If X_j reduces to an empty set, then delete the distribution symbol $p_j(X_j)$.

(b) (SIMPLIFICATION) Delete $p_j(X_j)$ and $p_h(V_h)$ if $X_j = V_h$.

If the algorithm reduces $f(X)$ to $p_i(X_i)$, we say that it terminates *with success.*

The functional expression

$$f(X) = \prod_i p_i(X_i) \Big/ \prod_h p_h(V_h)$$

is called a *product form* over the scheme $S = \{X_1, \ldots, X_s\}$, if it satisfies the set of marginal constraints $f(X_i) = p_i(X_i)$ for all X_i. If this is the case, each system of collectively compatible distributions over S admits an extension that has $f(X)$ as its functional expression. Such extensions will be referred to as *product extensions.*

We note that the functional expression of Example 4 is not a product form since it does not satisfy the marginal constraint

$$f(ABC) = p_1(ABC).$$

However, the following functional expression defined over the same scheme system

$$f(ABCDE) = \frac{p_1(ABC)p_2(AD)p_3(BE)}{p_2(A)p_3(B)}$$

is a product form.

The following theorem states a fundamental property of product extensions.

Theorem 8. *Product extensions maximize Shannon's entropy*

Proof. Let $f(x)$ and $p(x)$ be respectively a product extension and any other extension of the distribution system $\{p_1(x_1), \ldots, p_s(x_s)\}$. Using the well-known information-theoretical inequality [7]

$$-\sum p(x) \log p(x) \leq -\sum p(x) \log f(x),$$

we have

$$H[p(x)] \leq -\sum p(x)\Big[\sum_i \log p_i(x_i) - \sum_h \log p_h(v_h)\Big].$$

As the p_i's and the p_h's are also marginals of p, the right side reduces to

$$\sum_i H[p_i(x_i)] - \sum_h H[p_h(v_h)],$$

which is exactly the entropy of $f(x)$. □

Theorem 9. *If the functional expression*

$$f(X) = \prod_i p_i(X_i) \Big/ \prod_h p_h(V_h)$$

is a product form, then

(i) *the V_h's are unique,*

(ii) *the number of* V_h*'s is less or equal to* $s-1$,
(iii) *each* V_h *is a subset of two or more* X_i*'s.*

Proof. The statement (i) is an immediate consequence of the uniqueness of maximum-entropy extension. The statement (ii) follows from the successful termination of the simplification procedure. As to (iii), for the simplification procedure to terminate with success it is necessary that each V_h be a subset of some X_i. Now, assume that (iii) is false. Then, there is a V_{h^*} that is subset of only one scheme X_{i^*}. It is evident that in such a case the selective simplification procedure cannot terminate with success, since the distribution symbol $p_{h^*}(V_{h^*})$ cannot be deleted. □

The following theorem states that acyclicity is a necessary and sufficient condition for the existence of product forms.

Theorem 10. *A scheme system admits a product form if and only if it is acyclic.*

Proof. (if) Let $S=\{X_1,\ldots,X_s\}$ an acyclic scheme and $\{V_h\}$ be the set of its reducing factors. Then, the functional expression associated with S,

$$f(X)=\prod_i p_i(X_i)\Big/\prod_h p_h(V_h),$$

is a product form, that is, it satisfies each marginal constraint $f(X_i)=p_i(X_i)$. To see it, it is sufficient to run the selective simplification procedure according to a p.r.o. whose last element is X_i.

(only if) It suffices to prove that if $p_i(X_i)$ is deleted by simplification, then also X_i can be deleted by reduction. Now, this is an immediate consequence of the fact that by Theorem 9 the V_h's of a product form are subsets of two or more schemes X_i's. □

6. Collective compatibility

A general way to test the compatibility of given distributions is solving an equivalent problem of linear programming, whose variables are the values of a hypothetical distribution defined in the Cartesian product of $\Omega_1,\ldots,\Omega_s$:

$$\boldsymbol{\Omega}^*=\{(x_1,\ldots,x_s): x_1\in\Omega_1,\ldots,x_s\in\Omega_s\}.$$

A vector $(x_1,\ldots,x_s)$ in $\boldsymbol{\Omega}^*$ is said to be *consistent* if its components, $x_1,\ldots,x_s$, interpreted as singleton relations over $X_1,\ldots,X_s$, are collectively consistent. If this is the case and x is their extension to $X=\bigcup_i X_i$, then $(x_1,\ldots,x_s)$ is called the *representative* vector of x. The linear-programming problem consists in finding the solution, $\pi(x_1, \ldots, x_s)$, of the algebraic system

$$\begin{cases} \sum \pi(x_1, \ldots, x_s) = p_i(x_i) \quad (i = 1, \ldots, s), \\ \pi(x_1, \ldots, x_s) \geq 0, \end{cases}$$

that minimizes the "infeasibility form"

$$w = \sum c(x_1, \ldots, x_s)\pi(x_1, \ldots, x_s),$$

whose c-coefficients are defined as follows

$$c(x_1, \ldots, x_s) = \begin{cases} 0, & \text{if } (x_1, \ldots, x_s) \text{ is consistent,} \\ 1, & \text{else.} \end{cases}$$

Note that w measures the probability of the subset of Ω^* formed by all inconsistent vectors. Now, let $\pi(x_1, \ldots, x_s)$ be the solution computed by the symplex algorithm. If the minimum value of w is zero, then all inconsistent vectors have a zero probability. So π defines a proper extension of $p_1(x_1), \ldots, p_s(x_s)$:

$$p(x) = \pi(x_1, \ldots, x_s),$$

where $(x_1, \ldots, x_s)$ is the representative vector of x. But, if the minimum of w is positive, then no extensions exist. However, it is not always necessary to resort to the symplex algorithm for determining the compatibility of given distributions. For example, in case of partitions compatibility is always out of the question.

Theorem 11. *Let $S = \{X_1, \ldots, X_s\}$ be a scheme system and $X = \bigcup_i X_i$. The following conditions are equivalent*:

(i) *S is a partition of X,*

(ii) *there exists a product extension of each system $\{p_1(x_1), \ldots, p_s(x_s)\}$ of distributions over S,*

(iii) *there exists a common extension of each system $\{p_1(x_1), \ldots, p_s(x_s)\}$ of distributions over S.*

Proof. The implication "if (i), then (ii)" follows from Theorem 2.

The implication "if (ii), then (iii)" is trivial.

Finally, in order to prove the implication "if (iii), then (i)", by way of contradiction assume that it is false. Then, there exists a scheme system $S = \{X_1, \ldots, X_s\}$ that enjoys the property (iii) even if it is not a partition. But, if $S = \{X_1, \ldots, X_s\}$ is not a partition then in Ω^* there is a vector $(x_1^*, \ldots, x_s^*)$ that does not correspond to any X-tuple, that is, an inconsistent vector. Define the following distributions with schemes $X_1, \ldots, X_s$:

$$p_i(x_i) = \begin{cases} 1, & \text{if } x_i \text{ is equal to } x_i^*, \\ 0, & \text{else.} \end{cases}$$

Then, from the property (iii) it follows that the characteristic relations of $p_1(x_1), \ldots, p_s(x_s)$, that is, the vector $(x_1^*, \ldots, x_s^*)$, is collectively consistent. However, we showed that $(x_1^*, \ldots, x_s^*)$ is not consistent. This contradiction is the proof that the property (iii) holds for partitions only. □

At this point, we are able to answer the original question of compatibility. The following theorem states that the pairwise compatibility of given distributions is a necessary and sufficient condition for their collective compatibility if and only if the system of the distribution scheme is acyclic. Here the key point is that a product form is made up of the distributions in question and a certain set of their marginals. Therefore, for a product form to represent an extension it is sufficient that the distributions in question agree on those marginals, that is, be pairwise compatible.

Theorem 12. *Let $S = \{X_1, \ldots, X_s\}$ be a scheme system and $X = \bigcup_i X_i$. The following conditions are equivalent:*

(i) *S is acyclic,*

(ii) *there exists a product extension of each system $\{p_1(x_1), \ldots, p_s(x_s)\}$ of pairwise compatible distributions over S,*

(iii) *there exists a common extension of each system $\{p_1(x_1), \ldots, p_s(x_s)\}$ of pairwise compatible distributions over S.*

Proof. The implication "if (i), then (ii)" follows from the observation that Theorem 7 continues to hold even if the hypothesis of collective compatibility is replaced by that of pairwise compatibility, which is enough to make the product form (5) an extension of the marginals given.

The implication "if (ii), then (iii)" is trivial.

Finally, the implication "if (iii), then (i) is proved by contradiction. Assume there exists a cyclic scheme system $S = \{X_1, \ldots, X_s\}$ such that any system of pairwise compatible distributions with schemes $X_1, \ldots, X_s$ admit a common extension. Now, since the scheme system is cyclic, by virtue of the above-mentioned Relational Consistency Property there exists a system of relations $R_1, \ldots, R_s$, that are pairwise consistent but not collectively consistent. On the other hand, if $\{p_1(x_1), \ldots, p_s(x_s)\}$ is any system of pairwise compatible distributions having $R_1, \ldots, R_s$ as their characteristic relations, then there exists a common extension, which implies $R_1, \ldots, R_s$ are collectively consistent. This contradiction shows that the sufficiency of pairwise compatibility for collective compatibility does not apply to any cyclic scheme system. □

References

[1] C. Beeri, R. Fagin, D. Maier and M. Yannakakis, On the desirability of acyclic database schemes, J. Assoc. Comput. Mach. 30 (1983) 479–513.

[2] D.T. Brown, A note on approximations to discrete probability distributions, Inform. and Control 2 (1959) 386–392.
[3] N. Goodman and O. Shmueli, Syntactic characterization of tree database schemas, J. Assoc. Comput. Mach. 30 (1983) 767–786.
[4] H.H. Ku and S. Kullback, Approximations to discrete probability distributions, IEEE Trans. Inform. Theory 15 (1969) 444–447.
[5] P.M. Lewis, Approximating probability distributions to reduce storage requirements, Inform. and Control 2 (1959) 214–225.
[6] E. Marczewski, Measures in almost independent fields, Fund. Math. 38 (1951) 217–229.
[7] A. Rényi, Foundations of Probability (Holden-Day, San Francisco, 1970).
[8] R.T. Tarjan and M. Yannakakis, Simple linear-time algorithm to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs, SIAM J. Comput. 13 (1984) 566–579.
[9] M. Vlach, Conditions for the existence of solutions of the three-dimensional planar transportation problem, Discrete Appl. Math. 13 (1986) 61–78.