KYBERNETIKA — VOLUME *50* (2014), NUMBER 3, PAGES 322–362

EQUIVALENCE OF COMPOSITIONAL EXPRESSIONS AND INDEPENDENCE RELATIONS IN COMPOSITIONAL MODELS

Francesco M. Malvestuto

We generalize Jiroušek's (*right*) *composition operator* in such a way that it can be applied to distribution functions with values in a "semifield", and introduce (parenthesized) *compositional expressions*, which in some sense generalize Jiroušek's "generating sequences" of compositional models. We say that two compositional expressions are *equivalent* if their evaluations always produce the same results whenever they are defined. Our first result is that a set system $\mathcal{H}$ is star-like with centre X *if and only if* every two compositional expressions with "base scheme" $\mathcal{H}$ and "key" X are equivalent. This result is stronger than Jiroušek's result which states that, if $\mathcal{H}$ is star-like with centre X, then every two generating sequences with base scheme $\mathcal{H}$ and key X are equivalent. Then, we focus on *canonical expressions*, by which we mean compositional expressions θ such that the sequence of the sets featured in θ and arranged in order of appearance enjoys the "running intersection property". Since every compositional expression, whose base scheme is a star-like set system with centre X and whose key is X, is a canonical expression, we investigate the equivalence between two canonical expressions with the same base scheme and the same key. We state a graphical characterization of those set systems $\mathcal{H}$ such that every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent, and also provide a graphical algorithm for their recognition. Finally, we discuss the problem of detecting conditional independences that hold in a compositional model.

Keywords: compositional expression, compositional model, running intersection property, perfect sequence

Classification: 05C65, 05C85, 1699, 65C50, 60E99, 68T37

1. INTRODUCTION

Data pooling is a common practice in statistics [24] and consists in putting together "data from multiple data sources relating the same or different populations in order to obtain more precise estimates of common measurements of statistical information" [30]. In this spirit, probability distributions can be pooled to obtain a higher-dimensional probability distribution and, to achieve this, in a series of papers [7, 8, 11] Jiroušek introduced a binary operator "$\triangleright$", called (*right*) *composition*. Moreover, he proved that compositional models represent an alternative formalism to graphical models which are

DOI: 10.14736/kyb-2014-3-0322

used to model Bayesian networks, and turn out to be useful also in the framework of belief functions [12, 14, 15], possibility functions [14] and Shenoy valuations [13].

In the framework of probability distributions, Jiroušek [11] proved that

> there are many 'special situations' under which the order of application of the composition operator can be changed without influencing the resulting composed distribution.

These special situations are reported in Table 8 in [11] and some of them require the consistency of the system of the input probability distributions. In the spirit of data pooling and of "knowledge integration" [31], throughout we do not assume consistency so that the following result [11] is relevant to the object of this paper.

Theorem 1.1. Let f_1, f_2 and f_3 be probability distributions on X_1, X_2 and X_3, respectively. If $X_2 \cap X_3 \subseteq X_1$, then

$$f_1 \rhd (f_2 \rhd f_3) = f_1 \rhd (f_3 \rhd f_2) = (f_1 \rhd f_2) \rhd f_3 = (f_1 \rhd f_3) \rhd f_2.$$

In this paper, we first show how to apply the composition operator to arbitrary additive multivariate functions (of discrete variables) which take their values in a "semifield" (see Section 2 for basic definitions). We call them *distribution functions* and emphasize that their class includes not only probability distributions, but more in general multivariate functions whose values can be added, multiplied and divided, such as "relations" in databases [1] and "measures" in data warehouses [19, 23, 26, 27]. Then, we introduce the notion of a *compositional expression*, by which we mean a parenthesized expression formed out by *distinct* sets of variables, and the symbol $\rhd$. Structural elements of a compositional expression are its *base scheme*, its *base sequence* and its *key*. For example, the compositional expression $(AB \rhd CD) \rhd (BC \rhd AD)$ has the set system $\{AB, AD, BC, CD\}$ as its base scheme, the set sequence (AB, CD, BC, AD) as its base sequence and the set AB as its key. Next, we define the notion of *equivalence* between compositional expressions having the same base scheme and the same key. (Note that our notion of equivalence is stronger than that studied in [16, 17].) Thus, Theorem 1.1 can be re-stated as follows: If $X_2 \cap X_3 \subseteq X_1$, then the four compositional expressions

$$X_1 \rhd (X_2 \rhd X_3) \qquad X_1 \rhd (X_3 \rhd X_2) \qquad (X_1 \rhd X_2) \rhd X_3 \qquad (X_1 \rhd X_3) \rhd X_2$$

are pairwise equivalent. It should be noted that the four compositional expressions above are *simple* in that they contain exactly one subexpression of the form $(X \rhd Y)$.

After Jiroušek we call a set system $\mathcal{H}$ *star-like with centre* X if $X \in \mathcal{H}$ and $Y \cap Z \subseteq X$ for every two distinct sets Y and Z in $\mathcal{H}$. Accordingly, in its generalized form Theorem 1.1 reads:

> Given a star-like set system $\mathcal{H}$ with centre X, every two *simple* compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent.

We shall prove the following stronger result (see Theorem 6.5):

> A set system $\mathcal{H}$ is star-like with centre X *if and only if* every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent.

Next we focus on *canonical expressions*, a canonical expression being a compositional expression whose base sequence enjoys the *running intersection property* [1, 18]. Of course, if $\mathcal{H}$ is a star-like set system with centre X, then every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent. We then investigate the class of those set systems $\mathcal{H}$ for which every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent. We call such set systems X-*centric*, and provide a graphical recognition algorithm. As a result, the class of X-centric set systems strictly includes the class of star-like set systems with centre X.

Finally, we discuss problem of detecting conditional independences in the model generated by a compositional expression. For generating sequences of probability distributions this problem was discussed in [9, 10, 16, 17].

The paper is organized as follows. In Section 2 we give a precise definition of what we mean by a distribution function. In Section 3 we introduce the notion of conditional independence in a distribution function and state some properties. In Section 4 we introduce our composition operator which generalizes Jiroušek's composition operator to distribution functions. Section 5 is devoted to compositional expressions and we provide a general formula for the model generated by a compositional expression. In Section 6 we introduce the notion of equivalence between compositional expressions, and prove that star-like set systems with centre X are precisely those set systems $\mathcal{H}$ for which every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent. In Section 7 we introduce canonical expressions, and we provide a closed formula for the model generated by a canonical expression, which is used to characterize those set systems $\mathcal{H}$ such that every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent. Section 8 aims at finding out conditional independences holding in the model generated by a compositional expression. Finally, Section 9 contains a note on the power of the formalism of compositional expressions, as well as some directions for future research, and the Appendix contains the proof of the closed formula for canonical expressions.

2. PRELIMINARIES

2.1. Commutative semirings

A *commutative semiring* is a triple $(R, +, \times)$ where R is a set, and $+$ and $\times$ stand for operations such that

(P1) $(R, +, 0)$ is a commutative monoid, that is, the operation $+$ is associative and commutative, and there is an additive identity, denoted by 0, such that $a + 0 = a$ for all $a \in R$;

(P2) $(R, \times, 1)$ is a commutative monoid, that is, the operation $\times$ is associative and commutative, and there is a multiplicative identity, denoted by 1, such that $a \times 1 = a$ for all $a \in R$;

(P3) the distributive law holds, that is, $a \times (b + c) = (a \times b) + (a \times c)$ for all triples (a, b, c) from R.

Commutative semirings having the following two properties serve our purposes:

(P4) $(R, \times, 1)$ is a group, that is, for all $a \in R - \{0\}$ there is an element of R, denoted by a^{-1}, such that $a \times a^{-1} = 1$;

(P5) $a \times b = 0$ if and only if $a = 0$ or $b = 0$.

Such a commutative semiring will be referred to as a *semifield* since it is actually a field if every element of R admits an additive inverse. The following is a short list of semi-fields.

R	"$(+, 0)$"	"$(\times, 1)$"	short name
$(-\infty, +\infty)$	$(+, 0)$	$(\times, 1)$	real field
$[0, \infty)$	$(+, 0)$	$(\times, 1)$	sum-product semifield
$(0, \infty]$	$(\min, \infty)$	$(\times, 1)$	min-product semifield
$[0, \infty)$	$(\max, 0)$	$(\times, 1)$	max-product semifield
$(-\infty, +\infty]$	$(\min, +\infty)$	$(+, 0)$	min-sum semifield
$[-\infty, +\infty)$	$(\max, -\infty)$	$(+, 0)$	max-sum semifield
$\{0, 1\}$	$(+ \, mod \, 2, 0)$	$(\times, 1)$	Galois field GF(2)
$\{0, 1\}$	$(\vee, 0)$	$(\wedge, 1)$	Boolean algebra

Finally, observe that the sum-product, min-product, max-product, min-sum and max-sum semifields as well as Boolean algebra enjoy the property that the additive identity ("0") is the only element of R that has an additive inverse (which is equal to "0"), that is,

(P6) if $a + b = 0$ then $a = b = 0$.

Such semifields will be referred to as *metric semifields*. The simplest example of a non-metric semifield, to which we will refer, is the Galois field GF(2).

2.2. Distribution functions

Throughout we only consider discrete variables which take their values in finite sets and whose values are mutually exclusive and collectively exhaustive. We use the initial capital-case letters of the alphabet (e. g., A, B, C) to denote variables, and the other capital-case letters to denote sets of variables (e. g., X, Y, Z); moreover, sets of variables are written as lists of variables; thus, ABC stands for $\{A, B, C\}$. Let X be a set of variables. An X-*tuple* is an assignment of values to the variables in X; by $dom(X)$ we denote the set of all X-tuples. We use the lower-case letter x to denote an X-tuple. Let Y be a nonempty proper subset of X; given an X-tuple x, by x_Y we denote the Y-tuple obtained from x by ignoring the values of the variables in $X - Y$. It is convenient to introduce the following two operators of relational algebra [1].

Let r be a subset of $dom(X)$ and let Y be a nonempty subset of X. The *projection* of r onto Y, denoted by $\pi_Y(r)$, is the set of Y-tuples y for which there exists an X-tuple x in r such that $x_Y = y$:

$$\pi_Y(r) = \{x_Y : x \in r\}.$$

Let X and Y be two sets of variables, let $r \subseteq dom(X)$ and $s \subseteq dom(Y)$. The (*natural*) *join* of r and s, denoted by $r \bowtie s$, is the subset of $dom(X \cup Y)$ defined as follows:

$$r \bowtie s = \{z \in dom(X \cup Y) : z_X \in r \text{ and } z_Y \in s\}.$$

Note that if $X \cap Y = \emptyset$ then $r \bowtie s$ is a commutative form of the Cartesian product of r and s.

Remark 2.1. Let $r \subseteq dom(X)$ and let Z be a nonempty proper subset of X. Then $r \subseteq \pi_Z(r) \bowtie dom(X - Z)$.

Finally, it is easily seen that the join is an associative and commutative operator.

Let $(R, +, \times)$ be a semifield, and let X be a finite set of discrete variables. A *distribution function* on X is an R-valued function f on $dom(X)$, whose values can be added and multiplied according to the two operations of $(R, +, \times)$. If $(R, +, \times)$ is a metric semifield, then we call f a *metric distribution function* on X. Examples of metric distribution functions are distribution functions over the sum-product field (e. g., probability distributions) and over the min-product, max-product, min-sum and max-sum semifields and the Boolean algebra. Note that a distribution function over the Galois field GF(2) is not a metric distribution function.

A distribution function f on X is *uniform* if, for some $a \in R$, $f(x) = a$ everywhere (that is, for every X-tuple x); if $a = 1$ (respectively, $a = 0$), f is called a *unitary* (respectively, *null*) distribution function.

The *support* of a distribution function f on X, denoted by $\|f\|$, is the (possibly empty) set of X-tuples x with $f(x) \neq 0$. Note that $\|f\|$ uniquely determines f if $(R, +, \times)$ is the Boolean algebra or the Galois field GF(2). Let f and g be two distribution functions on X; we say that f is *dominated* by g, written $f \ll g$, if $\|f\| \subseteq \|g\|$.

Let Y be a nonempty proper subset of X; the *marginal* of f on Y, written $f^{\downarrow Y}$ using the Shenoy–Shafer notation [28], is the distribution function on Y defined as follows

$$f^{\downarrow Y}(y) = \sum_{x \in dom(X): x_Y = y} f(x)$$

where the summation symbol $\sum$ refers to the operation of addition (+) of the commutative semiring $(R, +, \times)$. Let $X - Y = \{A_1, \dots, A_k\}$. We can write an X-tuple x with $x_Y = y$ as $(a_1, \dots, a_k, y)$ where $a_h \in dom(A_h)$, $1 \leq h \leq k$. Then, one has

$$f^{\downarrow Y}(y) = \sum_{a_1 \in dom(A_1), \dots, a_k \in dom(A_k)} f(a_1, \dots, a_k, y).$$

Accordingly, in what follows, we also make use of the following sum-expression for $f^{\downarrow Y}$:

$$f^{\downarrow Y} = \sum_{A \in X - Y} f.$$

Finally, by $f^{\downarrow \emptyset}$ we denote the "norm" (or "grand-total") of f, that is, $f^{\downarrow \emptyset} = \sum_{x \in dom(X)} f(x)$.

Remark 2.2. For every Y-tuple y for which $f^{\downarrow Y}(y) \neq 0$, there always is an X-tuple x such that $x_Y = y$ and $f(x) \neq 0$. In other words, if $y \in \|f^{\downarrow Y}\|$ then there exists $x \in \|f\|$ such that $y = x_Y$ so that $y \in \pi_Y(\|f\|)$. To sum up, $\|f^{\downarrow Y}\| \subseteq \pi_Y(\|f\|)$.

Example 2.3. Let A and B be binary variables, and let f be the unitary distribution function on AB over the Galois field GF(2). Then $\|f\| = dom(AB)$, $\pi_A(\|f\|) = dom(A)$ and $\|f^{\downarrow A}\| = \emptyset$.

Let f be a distribution function on X. By an *extension* of f to a superset V of X we mean any distribution function e on V whose marginal on X coincides with f, that is, $e^{\downarrow X} = f$.

3. INDEPENDENCE RELATIONS

The probability-theoretic notion of conditional independence can be generalized in the framework of functions over a commutative semiring [21] and, hence, of our distribution functions. We first recall the definition and properties of probability-theoretic conditional independence, and then state properties of conditional independence in the framework of distribution functions.

3.1. Probability-theoretic conditional independence

Let X and Y be two disjoint nonempty sets of variables, and let Z be a (possibly empty) set of variables such that $Z \cap (X \cup Y) = \emptyset$. Let f be a probability distribution on $X \cup Y \cup Z$. The sets X and Y are *independent given* Z under f if for every $(X \cup Y \cup Z)$-tuple (x, y, z)

$$f(x, y, z) \times f^{\downarrow Z}(z) = f^{\downarrow X \cup Z}(x, z) \times f^{\downarrow Y \cup Z}(y, z). \qquad (1)$$

Let f be a probability distribution on a superset of $X \cup Y \cup Z$. Using Dawid's notation [4], we say that the *conditional independence* $X \perp\!\!\!\perp Y \mid Z$ *holds* in f if X and Y are independent given Z under $f^{\downarrow X \cup Y \cup Z}$. It is well-known that conditional independences satisfy the following properties, called *semigraphoid axioms* [18, 25].

(`symmetry axiom`)

if $X \perp\!\!\!\perp Y \mid Z$ holds in f, then $Y \perp\!\!\!\perp X \mid Z$ also holds in f;

(`decomposition axiom`)

if $X \perp\!\!\!\perp W \cup Y \mid Z$ holds under f , then $X \perp\!\!\!\perp Y \mid Z$ also holds in f;

(`weak-union axiom`)

if $X \perp\!\!\!\perp W \cup Y \mid Z$ holds in f, then $X \perp\!\!\!\perp Y \mid W \cup Z$ also holds in f;

(`contraction axiom`)

if $X \perp\!\!\!\perp Y \mid Z$ and $X \perp\!\!\!\perp W \mid Y \cup Z$ hold in f, then $X \perp\!\!\!\perp W \cup Y \mid Z$ also holds in f.

Finally, observe that eq. (1) can be re-written as

$$f(x, y, z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X \cup Z}(x,z) \times f^{\downarrow Y \cup Z}(y,z)}{f^{\downarrow Z}(z)} & \text{else.} \end{cases} \qquad (2)$$

3.2. Algebraic conditional independence

Consider now a distribution function f on a given semifield, and eqq. (1) and (2). Note that the factor $\frac{1}{f^{\downarrow Z}(z)}$ in eq. (2) does mean the multiplicative inverse of $f^{\downarrow Z}(z)$, that is, $\frac{1}{f^{\downarrow Z}(z)} = \left(f^{\downarrow Z}(z)\right)^{-1}$. The following example shows that eq. (1) and eq. (2) are not equivalent.

Example 3.1. Let A, B and C be binary variables with $dom(A) = \{a, \bar{a}\}$, $dom(B) = \{b, \bar{b}\}$ and $dom(C) = \{c, \bar{c}\}$, and let f be the unitary distribution function on ABC over the Galois field GF(2). Note that the marginal of f on every proper subset of ABC is a null distribution function. Therefore, eq. (1) holds everywhere (we always have the equality $1 \times 0 = 0 \times 0$), but, eq. (2) does not hold since $f^{\downarrow C}(c) = 0$ and $f(a, b, c) = 1$.

From Example 3.1 we learn that the notions underlying eq. (1) and eq. (2) are to be distinguished, which we are going to do. Let X and Y be two disjoint nonempty sets of variables, and let Z be a (possibly empty) set of variables such that $Z \cap (X \cup Y) = \emptyset$. Let f be a distribution function on $X \cup Y \cup Z$. We say that

— the sets X and Y are (*algebraically*) *independent* [21] *given* Z under f if the equality in eq. (1) holds everywhere;

— f is *decomposable* by the set pair $\{X \cup Z, Y \cup Z\}$ if the equality in eq. (2) holds everywhere.

Let f be a probability distribution on a superset of $X \cup Y \cup Z$. Again, we say that the *conditional independence* $X \perp\!\!\!\perp Y \mid Z$ *holds* in f if X and Y are independent given Z under $f^{\downarrow X \cup Y \cup Z}$.

First of all, we prove that both conditional independence and decomposability satisfy the symmetry axiom and the decomposition axiom.

Theorem 3.2. Conditional independence and decomposability of distribution functions satisfy the symmetry and decomposition axioms.

Proof. The symmetry axiom is a consequence of the commutativity of the multiplication. As for the decomposition axiom, assume that

$$f(w, x, y, z) \times f^{\downarrow Z}(z) = f^{\downarrow X \cup Z}(x, z) \times f^{\downarrow W \cup Y \cup Z}(w, y, z).$$

Summarizing over $X \cup Y \cup Z$ we obtain

$$f(x, y, z) \times f^{\downarrow Z}(z) = f^{\downarrow X \cup Z}(x, z) \times f^{\downarrow Y \cup Z}(y, z).$$

Analogously, from

$$f(w, x, y, z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X \cup Z}(x,z) \times f^{\downarrow W \cup Y \cup Z}(w,y,z)}{f^{\downarrow Z}(z)} & \text{else} \end{cases}$$

we obtain

$$f(x,y,z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X\cup Z}(x,z) \times f^{\downarrow Y\cup Z}(y,z)}{f^{\downarrow Z}(z)} & \text{else.} \end{cases}$$

□

However, in general neither conditional independence nor decomposability satisfies the weak-union axiom as is shown by the following two examples.

Example 3.3. Let A, B, C and D be binary variables with $dom(A) = \{a, \bar{a}\}$, $dom(B) = \{b, \bar{b}\}$, $dom(C) = \{c, \bar{c}\}$ and $dom(D) = \{d, \bar{d}\}$, and let f be the distribution function on $ABCD$ over the Galois field GF(2) with support

$$\|f\| = \{(a, b, c, d), (a, \bar{b}, \bar{c}, d), (\bar{a}, b, \bar{c}, d), (\bar{a}, \bar{b}, c, d)\}.$$

Consider the two conditional independences $A \perp\!\!\!\perp BC \mid D$ and $A \perp\!\!\!\perp B \mid CD$. Since the marginals of f on D and AD are null distribution functions, $A \perp\!\!\!\perp BC \mid D$ holds in f. As for $A \perp\!\!\!\perp B \mid CD$, since $f^{\downarrow CD}(c, d) = 0$ and $f^{\downarrow ACD}(a, c, d) = f^{\downarrow BCD}(b, c, d) = 1$, the equality

$$f(a, b, c, d) \times f^{\downarrow CD}(c, d) = f^{\downarrow ACD}(a, c, d) \times f^{\downarrow BCD}(b, c, d)$$

is not valid, which proves that $A \perp\!\!\!\perp B \mid CD$ does not hold in f.

Example 3.4. Let A, B, C and D be the binary variables of Example 3.3, and let f be the distribution function on $ABCD$ over the Galois field GF(2) with support

$$\|f\| = \{(a, b, c, d), (a, b, \bar{c}, d), (a, \bar{b}, \bar{c}, d), (\bar{a}, \bar{b}, c, \bar{d})\}.$$

Consider the two set pairs $\{AD, BCD\}$ and $\{ACD, BCD\}$. Since $f^{\downarrow D}$ is a unitary distribution function and

$$\|f^{\downarrow AD}\| = \{(a, d), (\bar{a}, \bar{d})\} \qquad \|f^{\downarrow BCD}\| = \{(b, c, d), (b, \bar{c}, d), (\bar{b}, c, \bar{d}), (\bar{b}, \bar{c}, d)\},$$

f is decomposable by $\{AD, BCD\}$. As for $\{ACD, BCD\}$, since $f^{\downarrow CD}(\bar{c}, d) = 0$ and $f(a, b, \bar{c}, d) = 1$, f is not decomposable by $\{ACD, BCD\}$.

We now prove that decomposability implies conditional independence.

Theorem 3.5. Let X and Y be two disjoint nonempty sets of variables, and let Z be a (possibly empty) set of variables such that $Z \cap (X \cup Y) = \emptyset$. Let f be a distribution function on $X \cup Y \cup Z$. If f is decomposable by $\{X \cup Z, Y \cup Z\}$, then X and Y are independent given Z under f.

P r o o f. Assume that the equality in eq. (2) holds everywhere. We want to prove that the equality in eq. (1) holds for every $(X \cup Y \cup Z)$-tuple (x, y, z). Let us distinguish the following two cases.

Case 1: $f^{\downarrow Z}(z) \neq 0$. In this case, by property (P4) of semifields, $f^{\downarrow Z}(z)$ has a multiplicative inverse and, then, the equality in eq. (1) trivially follows from the equality in eq. (2).

Case 2: $f^{\downarrow Z}(z) = 0$. In this case, $f(x, y, z) = 0$ by eq. (2) and, hence, the left-hand side of eq. (1) is zero by property (P5) of semifields. We now prove that also the right-hand side of eq. (1) is zero. Consider the factor

$$f^{\downarrow X \cup Z}(x, z) = \sum_{y'} f(x, y', z).$$

on the right-hand side of eq. (1). Since $f^{\downarrow Z}(z) = 0$, by eq. (2) each term $f(x, y', z)$ of the sum is 0. Therefore, $f^{\downarrow X \cup Z}(x, z) = 0$ and, hence, the right-hand side of eq. (1) is zero by property (P5) of semifields.

□

However, for metric distribution functions (that is, if the underlying semifield enjoys property (P6)), we shall prove that decomposability and conditional independence are equivalent, and that conditional independence also satisfies the weak-union and contraction axioms. To this end, we need the following technical lemma.

Lemma 3.6. Let f be a metric distribution and let $f^{\downarrow X}$ and $f^{\downarrow Y}$ be marginals of f with $Y \subseteq X$. If $f^{\downarrow Y}(y) = 0$, then $f^{\downarrow X}(x) = 0$ for every X-tuple x with $x_Y = y$.

Proof. By property (P6) of metric semifields. □

Note that, by Lemma 3.6, if $f^{\downarrow X}(x) \neq 0$ then $f^{\downarrow Y}(x_Y) \neq 0$.

Theorem 3.7. Let X and Y be two disjoint nonempty sets of variables, and let Z be a (possibly empty) set of variables such that $Z \cap (X \cup Y) = \emptyset$. A metric distribution function f on $X \cup Y \cup Z$ is decomposable by $\{X \cup Z, Y \cup Z\}$ if and only if X and Y are independent given Z under f.

Proof. (*Only if*) By Theorem 3.5. (*If*) Assume that $X \perp\!\!\!\perp Y \mid Z$ holds in f. We need to prove that the equality in eq. (2) holds for every $(X \cup Y \cup Z)$-tuple (x, y, z). If $f^{\downarrow Z}(z) = 0$, then $f(x, y, z) = 0$ by Lemma 3.6 and, if $f^{\downarrow Z}(z) \neq 0$, then the equality in eq. (2) trivially follows from the equality in eq. (1). □

Theorem 3.8. Conditional independence in metric distribution functions satisfies the weak-union axiom.

Proof. By Theorem 3.7, it is the same as to prove that decomposability satisfies the weak-union axiom. Let f be a metric distribution function on $W \cup X \cup Y \cup Z$ and assume that the equality in

$$f(w, x, y, z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X \cup Z}(x,z) \times f^{\downarrow W \cup Y \cup Z}(w,y,z)}{f^{\downarrow Z}(z)} & \text{else} \end{cases}$$

holds everywhere. We need to prove that the equality in

$$f(w,x,y,z) = \begin{cases} 0 & \text{if } f^{\downarrow W\cup Z}(w,z) = 0 \\ \frac{f^{\downarrow W\cup X\cup Z}(w,x,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow W\cup Z}(w,z)} & \text{else} \end{cases}$$

holds everywhere. If $f^{\downarrow W\cup Z}(w,z) = 0$ then $f(w,x,y,z) = 0$ by Lemma 3.6. Assume that $f^{\downarrow W\cup Z}(w,z) \neq 0$. By Lemma 3.6, one also has $f^{\downarrow Z}(z) \neq 0$ so that by hypothesis

$$f(w,x,y,z) = \frac{f^{\downarrow X\cup Z}(x,z) \times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Z}(z)}$$

which entails

$$\frac{f^{\downarrow W\cup X\cup Z}(w,x,z)}{f^{\downarrow W\cup Z}(w,z)} = \frac{f^{\downarrow X\cup Z}(x,z)}{f^{\downarrow Z}(z)}.$$

Therefore,

$$\frac{f^{\downarrow W\cup X\cup Z}(w,x,z) \times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow W\cup Z}(w,z)}$$

$$= \frac{f^{\downarrow X\cup Z}(x,z) \times \times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Z}(z)} = f(w,x,y,z)$$

which proves the statement. □

Theorem 3.9. Conditional independence in metric distribution functions satisfies the contraction axiom.

Proof. By Theorem 3.7, it is the same as to prove that decomposability satisfies the contraction axiom. Let f be a metric distribution function on $W\cup X\cup Y\cup Z$ and assume that both

$$f^{\downarrow X\cup Y\cup Z}(x,y,z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X\cup Z}(x,z)\times f^{\downarrow Y\cup Z}(y,z)}{f^{\downarrow Z}(z)} & \text{else} \end{cases} \tag{3}$$

$$f(w,x,y,z) = \begin{cases} 0 & \text{if } f^{\downarrow Y\cup Z}(y,z) = 0 \\ \frac{f^{\downarrow X\cup Y\cup Z}(x,y,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Y\cup Z}(y,z)} & \text{else} \end{cases} \tag{4}$$

hold everywhere. We need to prove that the equality in

$$f(w,x,y,z) = \begin{cases} 0 & \text{if } f^{\downarrow Z}(z) = 0 \\ \frac{f^{\downarrow X\cup Z}(x,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Z}(z)} & \text{else} \end{cases}$$

holds everywhere.
If $f^{\downarrow Z}(z) = 0$ then $f(w,x,y,z) = 0$ by Lemma 3.6.
Assume that $f^{\downarrow Z}(z) \neq 0$. Let us distinguish two cases depending on whether or not $f^{\downarrow Y\cup Z}(y,z) = 0$.

Case 1: $f^{\downarrow Y\cup Z}(y,z)=0$. In this case, by Lemma 3.6 one has both $f(w,x,y,z)=0$ and $f^{\downarrow W\cup Y\cup Z}(w,y,z)=0$. Therefore, the equality

$$f(w,x,y,z)=\frac{f^{\downarrow X\cup Z}(x,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Z}(z)}$$

holds, which proves the statement.

Case 2: $f^{\downarrow Y\cup Z}(y,z)\neq 0$. Since $f^{\downarrow Z}(z)\neq 0$ and $f^{\downarrow Y\cup Z}(y,z)\neq 0$, by eq. (3) one has

$$\frac{f^{\downarrow X\cup Y\cup Z}(x,y,z)}{f^{\downarrow Y\cup Z}(y,z)}=\frac{f^{\downarrow X\cup Z}(x,z)}{f^{\downarrow Z}(z)}$$

so that by eq. (4) one also has

$$\begin{aligned}f(w,x,y,z)&=\frac{f^{\downarrow X\cup Y\cup Z}(x,y,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Y\cup Z}(y,z)}\\&=\frac{f^{\downarrow X\cup Z}(x,z)\times f^{\downarrow W\cup Y\cup Z}(w,y,z)}{f^{\downarrow Z}(z)}\end{aligned}$$

which proves the statement.

□

4. THE COMPOSITION OPERATOR

Let f and g be distribution functions on X and Y, respectively, over the same semifield. We say that f is *composable* with g, written $f\propto g$, if

(a) either $X\cap Y=\emptyset$ and $g^{\downarrow\emptyset}\neq 0$, or

(b) $X\cap Y\neq\emptyset$ and, for every X-tuple x, if $f(x)\neq 0$ then $g^{\downarrow X\cap Y}(x_{X\cap Y})\neq 0$.

Remark 4.1. Condition (b) requires that $X\cap Y\neq\emptyset$ and $\pi_{X\cap Y}(\|f\|)\subseteq\|g^{\downarrow X\cap Y}\|$.

First of all, we state some simple algebraic properties of composability. First of all, for $X=Y$ one has that $f\propto g$ if and only if $f\ll g$ so that composability turns out to be a reflexive relation (that is, $f\propto f$). Moreover, in general it is not symmetric (that is, $f\propto g$ does not imply $g\propto f$). Finally, it is not transitive (that is, $f\propto g$ and $g\propto h$ do not imply $f\propto h$). To see it, consider a distribution function f on X and two distribution functions g and h both on Y and assume that $X\cap Y=\emptyset$, $g^{\downarrow\emptyset}\neq 0$, $h^{\downarrow\emptyset}=0$ and $\|g\|=\|h\|$. Then, since $g^{\downarrow\emptyset}\neq 0$ and $h^{\downarrow\emptyset}=0$, $f\propto g$ holds but $f\propto h$ does not hold; on the other hand, $g\propto h$ holds since $\|g\|=\|h\|$.

Let f and g be distribution functions on X and Y, respectively. Consider the distribution function k on $V=X\cup Y$ defined as follows:

- if $X\cap Y=\emptyset$ then $k(v)=f(v_X)\times\frac{g(v_Y)}{g^{\downarrow\emptyset}}$;

- if $X \cap Y \neq \emptyset$ then $k(v) = \begin{cases} 0 & \text{if } f(v_X) = 0 \\ f(v_X) \times \frac{g(v_Y)}{g^{\downarrow X\cap Y}(v_{X\cap Y})} & \text{else.} \end{cases}$

Then, it is easy to see that the distribution function k is well-defined if and only if f is composable with g. To sum up, we write

$$k = \begin{cases} f \times \frac{g}{g^{\downarrow X\cap Y}} & \text{if } f \propto g \\ \text{undefined} & \text{else.} \end{cases}$$

Remark 4.2. If k is defined then $\|k\| = \|f\| \bowtie \|g\|$ and $\pi_X(\|k\|) = \|f\|$.

The following result is obvious.

Fact 4.3. If f is composable with g then the distribution function $k = f \times \frac{g}{g^{\downarrow X\cap Y}}$ is an extension of f to $X \cup Y$.

Theorem 4.4. Let f and g be distribution functions on X and Y, respectively, over the same semifield, and assume that neither $X - Y$ nor $Y - X$ is the empty set. If f is composable with g, then the sets $X - Y$ and $Y - X$ are independent given $X \cap Y$ under $k = f \times \frac{g}{g^{\downarrow X\cap Y}}$.

Proof. By eq. (1), we need to prove that

$$k \times k^{\downarrow X\cap Y} = k^{\downarrow X} \times k^{\downarrow Y}. \tag{5}$$

By Fact 4.3, one has $k^{\downarrow X} = f$ and, hence, $k^{\downarrow X\cap Y} = f^{\downarrow X\cap Y}$. Therefore, the left-hand side and the right-hand side of (5) can be written as

$$f \times \frac{g}{g^{\downarrow X\cap Y}} \times f^{\downarrow X\cap Y} \qquad\qquad f \times k^{\downarrow Y}$$

respectively. Moreover, we can write the marginal of k on Y as follows

$$k^{\downarrow Y} = \sum_{A\in X-Y} k = \frac{g}{g^{\downarrow X\cap Y}} \times \sum_{A\in X-Y} f = \frac{g}{g^{\downarrow X\cap Y}} \times f^{\downarrow X\cap Y}$$

so that the left-hand and right-hand sides of (5) are equal. □

The following example shows that the sets $X - Y$ and $Y - X$ are independent given $X \cap Y$ under $k = f \times \frac{g}{g^{\downarrow X\cap Y}}$, but k is not decomposable by $\{X, Y\}$.

Example 4.5. Let A, B and C be binary variables with $dom(A) = \{a, \bar{a}\}$, $dom(B) = \{b, \bar{b}\}$ and $dom(C) = \{c, \bar{c}\}$. Consider the unitary distribution function f on AB and the distribution function g on BC with support $\|g\| = \{(b, c), (\bar{b}, c)\}$, both over the Galois field GF(2). First of all, observe that $g^{\downarrow B}$ is the unitary distribution on B; therefore,

f is composable with g and $k = f \times \frac{g}{g^{\downarrow B}} = f \times g$. The support of k is $\|k\| = \|f\| \bowtie \|g\| = \{(a,b,c),(a,\bar{b},c),(\bar{a},b,c),(\bar{a},\bar{b},c)\}$. Now, since $f^{\downarrow A}$ is the null distribution on A, one has that $k^{\downarrow BC}$ $(= f^{\downarrow A} \times g)$ is the null distribution on BC and, hence, $k^{\downarrow B}$ is the null distribution on B. Therefore, eq. (1) holds everywhere which proves that the conditional independence $A \perp\!\!\!\perp C \mid B$ holds in k. However, since $k^{\downarrow B}(b) = 0$ and $k(a,b,c) = 1$, k is not decomposable by $\{AB, BC\}$.

After Jirousek [7, 11] we use the (binary) *composition operator* "$\triangleright$" to denote the distribution function k and write:

$$f \triangleright g = \begin{cases} f \times \frac{g}{g^{\downarrow X\cap Y}} & \text{if } f \propto g \\ \text{undefined} & \text{else.} \end{cases} \tag{6}$$

It is worth pointing out that a composition-like operator (see the "fitting operator" [20]) is tacitly present in the Proportional Fitting (or Scaling) Procedure used in the statistical analysis of contingency tables [5].

It is easy to see that, in general, the composition operator is neither commutative nor associative, and is idempotent, that is, $f \triangleright f = f$ (for $X = Y$, $f \triangleright g$ is defined if and only if $f \ll g$ and, then, $f \triangleright g = f$).

4.1. Composition of metric distribution functions

In this subsection we first give some properties of the composition of metric distribution functions.

Remark 4.6. Let f be a metric distribution function on X, and let Y be a nonempty subset of X. For every X-tuple $x \in \|f\|$, the Y-tuple x_Y belongs to $\|f^{\downarrow Y}\|$ so that by Remark 2.2 one has that $\|f^{\downarrow Y}\| = \pi_Y(\|f\|)$.

Let f and g be metric distribution functions on X and Y, respectively. By Remarks 4.1 and 4.6, the conditions (a) and (b) that ensure that f is composable with g are equivalent to the following conditions (a') and (b'), respectively:

(a') either $X \cap Y = \emptyset$ and g is not the null distribution function on Y, or

(b') $X \cap Y \neq \emptyset$ and $f^{\downarrow X\cap Y} \ll g^{\downarrow X\cap Y}$.

Assume that f is composable with g. We shall prove that $f \triangleright g = f \triangleright h$ where h denotes the *trivial extension* of g to $V = X \cup Y$, by which we mean the (metric) distribution function on V defined as follows: if $X \subseteq Y$ then $h = g$; otherwise, for every V-tuple v one has

$$h(v) = \frac{g(v_Y)}{dom(X - Y)}.$$

Note that, if X is not a subset of Y then

$$h^{\downarrow X} = \sum_{A \in Y-X} \frac{g}{dom(X-Y)} = \frac{g^{\downarrow X\cap Y}}{dom(X-Y)}$$

so that

$$\frac{h}{h^{\downarrow X}} = \frac{g}{g^{\downarrow X \cap Y}}. \tag{7}$$

and

$$\|h^{\downarrow X}\| = \|g^{\downarrow X \cap Y}\| \bowtie dom(X - Y). \tag{8}$$

Lemma 4.7. Let f and g be metric distribution functions on X and Y, respectively, and let h be the trivial extension of g to $V = X \cup Y$. Then, $f \triangleright h = f \triangleright g$.

Proof. The statement is trivially true if $X \subseteq Y$ for, then, $h = g$. Assume that $X - Y \neq \emptyset$. Suppose initially that both $f \triangleright h$ and $f \triangleright g$ are defined. Then, by eq. (7), one has

$$f \triangleright h = f \times \frac{h}{h^{\downarrow X}} = f \times \frac{g}{g^{\downarrow X}} = f \triangleright g.$$

At this point, we need only to prove that $f \propto h$ if and only if $f \propto g$. If $X \cap Y = \emptyset$ then the statement easily follows from the above-mentioned condition (a'). Assume that $X \cap Y \neq \emptyset$. By the above-mentioned condition (b'), one has that $f \propto h$ if and only if $f \ll h^{\downarrow X}$ and that $f \propto g$ if and only if $f^{\downarrow X \cap Y} \ll g^{\downarrow X \cap Y}$. Therefore, we need to prove that $f \ll h^{\downarrow X}$ if and only if $f^{\downarrow X \cap Y} \ll g^{\downarrow X \cap Y}$.

(*If*) Assume that $\|f^{\downarrow X \cap Y}\| \subseteq \|g^{\downarrow X \cap Y}\|$. By Remarks 2.1 and 4.6 and by eq. (8) one has

$$\|f\| \subseteq \pi_{X \cap Y}(\|f\|) \bowtie dom(X - Y)$$

$$= \|f^{\downarrow X \cap Y}\| \bowtie dom(X - Y) \subseteq \|g^{\downarrow X \cap Y}\| \bowtie dom(X - Y) = \|h^{\downarrow X}\|.$$

(*Only if*) Assume that $\|f\| \subseteq \|h^{\downarrow X}\|$. By eq. (8) and Remark 4.6 one has

$$\|f\| \subseteq \|g^{\downarrow X \cap Y}\| \bowtie dom(X - Y) = \pi_{X \cap Y}(\|g\|) \bowtie dom(X - Y)$$

so that

$$\pi_{X \cap Y}(\|f\|) \subseteq \pi_{X \cap Y}\Big(\pi_{X \cap Y}(\|g\|) \bowtie dom(X - Y)\Big) = \pi_{X \cap Y}(\|g\|)$$

and again by Remark 4.6

$$\|f^{\downarrow X \cap Y}\| \subseteq \|g^{\downarrow X \cap Y}\|.$$

□

Finally, the following result states decomposability of the composition of metric distribution functions.

Theorem 4.8. Let f and g be metric distribution functions on X and Y, respectively, over the same (metric) semifield. If f is composable with g, then $f \triangleright g$ is decomposable by $\{X, Y\}$.

Proof. By Theorems 4.4 and 3.7. □

4.2. Composition of probability distributions

A *probability distribution* on X is a distribution function f on X over the sum-product semifield such that $f^{\downarrow\emptyset} = 1$. Since the sum-product semifield is metric, every probability distribution is a metric distribution function. We shall provide an information-theoretic characterization of the composition of two probability distributions.

Let f and g be probability distributions on X and Y, respectively. By conditions (a') and (b') in Subsection 4.1, f is composable with g if and only if either $X \cap Y = \emptyset$ or $f^{\downarrow X\cap Y} \ll g^{\downarrow X\cap Y}$, which are precisely the requirements used in [11, 12]. Assume that $f \propto g$. By Fact 4.3, $f \triangleright g$ is an extension of f to $X \cup Y$. We shall prove that $f \triangleright g$ is the extension of f to $X \cup Y$ that is "closest" to g in an information-theoretic sense. To this end, we first consider two probability distributions e and h on V such that $e \ll h$. It is well-known the *I-divergence* [3] of h from e (also called the "cross-entropy" of e with respect to h or the "Kullback–Leibler divergence" of h from e)

$$I(e,h) = \sum_{v\in\|e\|} e(v)\ \log \frac{e(v)}{h(v)}$$

is a nonnegative quantity which vanishes if and only if $e = h$. Consider now a probability distribution f on X, and a probability distribution h on a superset V of X such that $f \ll h^{\downarrow X}$. Jiroušek (see Theorem 6.2 in [11]) proved the following information-theoretic characterization of $f \triangleright h$.

Theorem 4.9. For every extension e of f to V, one has $I(e,h) = I(f\triangleright h, h) + I(e, f\triangleright h)$.

Since $I(e, f \triangleright h) \geq 0$, by Theorem 4.9 one has that $I(f \triangleright h, h) \leq I(e,h)$ for every extension e of f to V; accordingly, $f \triangleright h$ is called the *I-projection* of h onto the set of extensions of f to V [11]. Finally, let f be a probability distribution on X, and let g be a probability distribution on Y such that $f \propto g$. Let h be the trivial extension of g to $V = X \cup Y$. By Theorem 4.9, $f \triangleright h$ is the I-projection of h onto the set of extensions of f to V and, by Lemma 4.7, $f \triangleright h = f \triangleright g$. Therefore, $f \triangleright g$ is the I-projection of the trivial extension of g to V onto the set of extensions of f to V; in this sense we can say that $f \triangleright g$ is the extension of f that is "closest" to g.

5. COMPOSITIONAL EXPRESSIONS

A *compositional expression* is a parenthesized expression formed out by *distinct* non-empty sets of variables, and the symbol "$\triangleright$". Explicitly, the following provides a formal definition of a compositional expression:

(i) if X is a set of variables, then X is a compositional expression;

(ii) if θ_1 and θ_2 are compositional expressions and no set in θ_2 occurs in θ_1, then $(\theta_1) \triangleright (\theta_2)$ is a compositional expression.

Given a compositional expression θ, by α_θ we denote the sequence of the sets featured in θ arranged according to the order of appearance; we call α_θ the *base sequence* of θ. Let $\alpha_\theta = (X_1, X_2, \ldots, X_n)$, $n \geq 1$. We call the set X_1 the *key* of θ; moreover, if $n > 1$, for each $i > 1$ by $\partial_\alpha X_i$ we denote the set $(\cup_{1\le j\le i-1} X_j) \cap X_i$. We also use the following notation:

$$V(\theta) = \cup_{1\le i\le n} X_i \quad \mathrm{COM}(\theta) = \cup_{i\ne j} X_i \cap X_j \quad \mathrm{UNI}(\theta) = V(\theta) - \mathrm{COM}(\theta);$$

thus, $\mathrm{COM}(\theta)$ is the set of variables that are common to at least two distinct sets in θ.

Henceforth, a compositional expression of either form $(X) \rhd (\theta)$ or $(\theta) \rhd (X)$ or $(X) \rhd (Y)$ will be written simply as $X \rhd (\theta)$ or $(\theta) \rhd X$ or $X \rhd Y$, respectively.

A *subexpression* of a compositional expression θ is defined as usual. Explicitly, a compositional expression θ' is a subexpression of θ if θ' is a substring of θ. A subexpression θ' of θ is *atomic* if it is of the form $\theta' = X$. We shall prove (see Theorem 5.4 below) that the number of subexpressions of a compositional expression formed out by n sets is equal to $2n - 1$.

The syntactic structure of a compositional expression θ can be represented by an (ordered full) binary tree T, to be called the *syntax tree* for θ, whose leaves correspond one-to-one to the atomic subexpressions of θ, and whose interior nodes correspond one-to-one to the non-atomic subexpressions of θ; explicitly, an interior node v of T corresponds to the subexpression $\theta' = (\theta_1) \rhd (\theta_2)$ of θ if θ_1 is the subexpression of θ corresponding to the "first" child of v, and θ_2 is the subexpression of θ corresponding to the "second" child of v.

5.1. Compositional model

Let θ be a compositional expression with $\alpha_\theta = (X_1, X_2, \ldots, X_n)$. A sequence $\mathbf{f} = (f_1, \ldots, f_n)$ in which f_i is a distribution function on X_i, $1 \le i \le n$, is called a (*functional*) *interpretation* of θ. Henceforth, we assume that the distribution functions $f_1, \ldots, f_n$ are all over the same semifield.

Let θ' be any subexpression of θ and let $\alpha_{\theta'} = (X_k, \ldots, X_m)$ be the base sequence of θ', for some k and m, $1 \le k \le m \le n$. By $[\theta']_{\mathbf{f}}$ we denote the result of replacing each set X_i, $k \le i \le m$, with the distribution function f_i, and then applying the composition operator if θ' is a non-atomic subexpression (that is, if $m > k$). We say that $\mathbf{f}$ is a *valid interpretation* of θ if $[\theta]_{\mathbf{f}}$ is defined. It should be noted that $\mathbf{f}$ is a valid interpretation of θ if and only if, for every subexpression $(\theta') \rhd (\theta'')$ of θ, both $[\theta']_{\mathbf{f}}$ and $[\theta'']_{\mathbf{f}}$ are defined and $[\theta']_{\mathbf{f}} \propto [\theta'']_{\mathbf{f}}$. If $\mathbf{f}$ is a valid interpretation of θ then, by Fact 4.3, $[\theta]_{\mathbf{f}}$ is an extension to $V(\theta)$ of the distribution function (f_1) on the key (X_1) of θ. We call $[\theta]_{\mathbf{f}}$ the *value* of θ under $\mathbf{f}$. From a computational point of view, $[\theta]_{\mathbf{f}}$ can be obtained with a bottom-up traversal of the syntax tree T for θ. Initially, each leaf of T is charged with the corresponding distribution function in $\mathbf{f}$; then, when an interior node v is examined, if u and w are the first child and the second child of v respectively, the node v is charged with the distribution function $g \rhd h$ where g and h are the charges on u and w, respectively. Finally, the charge on the root of T provides $[\theta]_{\mathbf{f}}$.

The *model generated* by a compositional expression θ, denoted by M_θ, is the set of the distribution functions $[\theta]_{\mathbf{f}}$ for all valid interpretations $\mathbf{f}$ of θ.

We shall provide a general formula for the value of a compositional expression under a valid interpretation. To achieve this, we need the following lemma.

Lemma 5.1. Let θ be a compositional expression, and let θ' be any subexpression of θ. Let $\alpha_\theta = (X_1, X_2, \ldots, X_n)$ and $\alpha_{\theta'} = (X_k, \ldots, X_m)$. Given a valid interpretation $\mathbf{f} = (f_1, \ldots, f_n)$ of θ, one has

$$[\theta']_{\mathbf{f}} = \frac{1}{q_{\theta'}} \times \prod_{k \le i \le m} f_i \tag{9}$$

where $q_{\theta'}$ is a function of $\mathrm{COM}(\theta')$ if $\mathrm{COM}(\theta') \neq \emptyset$, and is a constant otherwise.

Proof. We prove the statement by induction on the number of sets in $\alpha_{\theta'}$.
BASIS. If $k = m$ then $\mathrm{COM}(\theta') = \emptyset$ and eq. (9) holds with $q_{\theta'} = 1$.
INDUCTION. Assume that $k < m$ and let $\theta' = (\theta_1) \rhd (\theta_2)$. Let $\alpha_{\theta_1} = (X_k, \ldots, X_l)$ and $\alpha_{\theta_2} = (X_{l+1}, \ldots, X_m)$ for some l, $k \le l \le m-1$. By the inductive hypothesis, one has

$$[\theta_1]_{\mathbf{f}} = \frac{1}{q_{\theta_1}} \times \prod_{k \le i \le l} f_i \qquad [\theta_2]_{\mathbf{f}} = \frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i$$

where q_{θ_h} is a function of $\mathrm{COM}(\theta_h)$ if $\mathrm{COM}(\theta_h) \neq \emptyset$, and is a constant otherwise, $h = 1, 2$. By eq. (6) one has

$$[\theta']_{\mathbf{f}} = \frac{1}{q_{\theta_1}} \times \prod_{k \le i \le l} f_i \times \frac{\frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i}{\sum_{A \in V(\theta_2) - V(\theta_1)} \left(\frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i \right)},$$

which with

$$q_{\theta'} = q_{\theta_1} \times q_{\theta_2} \times \sum_{A \in V(\theta_2) - V(\theta_1)} \left(\frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i \right) \tag{10}$$

reduces to eq. (9). At this point, we only need to prove that $q_{\theta'}$ is a function of $\mathrm{COM}(\theta')$ if $\mathrm{COM}(\theta') \neq \emptyset$, and is a constant otherwise. Let us distinguish the following two cases:

Case 1: $V(\theta_1) \cap V(\theta_2) = \emptyset$. In this case, $\sum_{A \in V(\theta_2) - V(\theta_1)} \left(\frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i \right)$ is a constant and, since $\mathrm{COM}(\theta') = \mathrm{COM}(\theta_1) \cup \mathrm{COM}(\theta_2)$, $q_{\theta'}$ is a function of $\mathrm{COM}(\theta')$ if $\mathrm{COM}(\theta_1) \neq \emptyset$ or $\mathrm{COM}(\theta_2) \neq \emptyset$, and is a constant otherwise.

Case 2: $V(\theta_1) \cap V(\theta_2) \neq \emptyset$. In this case, $\sum_{A \in V(\theta_2) - V(\theta_1)} \left(\frac{1}{q_{\theta_2}} \times \prod_{l+1 \le i \le m} f_i \right)$ is a function of $V(\theta_1) \cap V(\theta_2)$ so that, since $\mathrm{COM}(\theta') = \mathrm{COM}(\theta_1) \cup \mathrm{COM}(\theta_2) \cup (V(\theta_1) \cap V(\theta_2))$, $q_{\theta'}$ is a function of $\mathrm{COM}(\theta')$.

□

Theorem 5.2. Let θ be a compositional expression, and let $\mathbf{f} = (f_1, \ldots, f_n)$ be a valid interpretation of θ. Then one has

$$[\theta]_{\mathbf{f}} = \frac{1}{q_\theta} \times \prod_{1 \le i \le n} f_i \tag{11}$$

where q_θ is a function of $\mathrm{COM}(\theta)$ if $\mathrm{COM}(\theta) \neq \emptyset$, and is a constant otherwise.

Proof. For $\theta' = \theta$, eq. (9) reduces to eq. (11). □

The following is an illustrative example.

Example 5.3. Consider the following compositional expression

$$\theta = (ABE \triangleright CF) \triangleright (CDG \triangleright ADH).$$

Then we have $\alpha_\theta = (ABE, CF, CDG, ADH)$, $\mathrm{COM}(\theta) = ACD$ and $\mathrm{UNI}(\theta) = BEFGH$. Consider now the following two subexpressions of θ:

$$\theta_1 = ABE \triangleright CF \qquad \theta_2 = CDG \triangleright ADH.$$

Then, one has

h	$V(\theta_h)$	$\mathrm{COM}(\theta_h)$	$\mathrm{UNI}(\theta_h)$
1	$ABCEF$	$\emptyset$	$ABCEF$
2	$ACDGH$	D	$ACGH$

Let $\mathbf{f} = (f_1, \ldots, f_4)$ be an interpretation of θ. Then, $\mathbf{f}$ is a valid interpretation of θ if

— $[\theta_1]_\mathbf{f} = f_1 \triangleright f_2$ is defined, that is, if $f_1 \propto f_2$,

— $[\theta_2]_\mathbf{f} = f_3 \triangleright f_4$ is defined, that is, if $f_3 \propto f_4$, and

— $[\theta_1]_\mathbf{f} \triangleright [\theta_2]_\mathbf{f}$ is defined, that is, if $(f_1 \triangleright f_2) \propto (f_3 \triangleright f_4)$.

If this is the case, then

— $[\theta_1]_\mathbf{f} = \frac{f_1 \times f_2}{f_2^{\downarrow\emptyset}}$ which reduces to eq. (9) with $q_{\theta_1} = f_2^{\downarrow\emptyset}$, which is a constant according to $\mathrm{COM}(\theta_1) = \emptyset$;

— $[\theta_2]_\mathbf{f} = \frac{f_3 \times f_4}{f_4^{\downarrow D}}$ which reduces to eq. (9) with $q_{\theta_2} = f_4^{\downarrow D}$, which is a function of $\mathrm{COM}(\theta_2) = D$.

Finally, we have

$$\begin{aligned} [\theta]_\mathbf{f} &= \frac{f_1 \times f_2}{f_2^{\downarrow\emptyset}} \times \frac{\frac{f_3 \times f_4}{f_4^{\downarrow D}}}{\sum_{D,G,H}\left(\frac{f_3 \times f_4}{f_4^{\downarrow D}}\right)} \\ &= \frac{f_1 \times f_2 \times f_3 \times f_4}{f_2^{\downarrow\emptyset} \times f_4^{\downarrow D} \times \sum_D \frac{f_3^{\downarrow CD} \times f_4^{\downarrow AD}}{f_4^{\downarrow D}}} \end{aligned}$$

which reduces to eq. (11) with

$$q_\theta = f_2^{\downarrow\emptyset} \times f_4^{\downarrow D} \times \sum_D \frac{f_3^{\downarrow CD} \times f_4^{\downarrow AD}}{f_4^{\downarrow D}}$$

which is a function of $\mathrm{COM}(\theta) = ACD$.

5.2. Sequential compositional expressions

We now introduce a type of compositional expression of special interest. Recall from the Introduction that a compositional expression is simple if it has exactly one subexpression of the type $X \triangleright Y$. If θ is a simple compositional expression and $\mathbf{f}$ is a valid interpretation of θ, then during the process of evaluation of θ under $\mathbf{f}$ only one "intermediate table" is to be maintained at a time. (Note that the compositional expressions involved in Theorem 1.1 are all simple.) Special simple compositional expressions are compositional expressions of the form

$$\theta = \Big(\big(\ldots(X_1 \triangleright X_2) \triangleright \ldots\big) \triangleright X_{n-1}\Big) \triangleright X_n \tag{12}$$

which we call *sequential compositional expressions* and correspond to Jiroušek's *generating sequences.* Note that a sequential compositional expression is uniquely determined by its base sequence; accordingly, as in [16] and [17] the base sequence of a sequential compositional expression may be called its "structure". A further property of a sequential compositional expression is that, by Fact 4.3, each intermediate table is always an extension of the distribution function f_1 in $\mathbf{f}$ on X_1 (which is the key of θ). Finally, it is easy to see that for the sequential compositional expression (12) the function q_θ in eq. (11) is simply a product of marginal distribution functions; explicitly, for $[\theta]_\mathbf{f}$ one has the following closed form:

$$[\theta]_\mathbf{f} = f_1 \times \prod_{2 \le i \le n} \frac{f_i}{f_i^{\downarrow \partial_\theta X_i}}. \tag{13}$$

Before closing this section, we observe that the sequential compositional expression (12) has exactly n atomic subexpressions and $n-1$ non-atomic subexpressions, each of which is of the form $\Big(\big(\ldots(X_1 \triangleright X_2) \triangleright \ldots\big) \triangleright X_{i-1}\Big) \triangleright X_i$, $2 \le i \le n$. Therefore, the number of the subexpressions of the sequential compositional expression (12) is $2n-1$. In other words, the syntax tree for the sequential compositional expression (12) has $2n-1$ nodes (n leaves plus $n-1$ interior nodes). Starting from this fact, we can prove the following more general result.

Theorem 5.4. The number of the subexpressions of a compositional expression formed out by n sets is $2n-1$.

P r o o f. Let θ be a compositional expression with $\alpha_\theta = (X_1, X_2, \ldots, X_n)$. It is sufficient to prove that the syntax tree T for θ has exactly $2n-1$ nodes. To achieve this, we show that T can be transformed into the syntax tree for the sequential compositional expression (12) without changing the number of nodes of T. Then, since the syntax tree for the sequential compositional expression (12) has exactly $2n-1$ nodes, we can conclude that also T has $2n-1$ nodes, which implies that the number of the subexpressions of θ is $2n-1$.

The transformation of T is carried out by processing the nodes (leaves) $v_n, v_{n-1}, \ldots, v_3$ of T corresponding to $X_n, X_{n-1}, \ldots, X_3$, respectively. Let a be the root of T.

Step 1. If v_n has distance greater than 1 from a, then

— add one node r_n to T which becomes the new root of T;

— make a the first child of r_n;

— let p be the parent of v_n; identify p with its first child and make v_n the second child of r_n.

Step 2. For $i = n-1, \ldots, 3$, do:

Let u be the sibling of v_{i+1}. If v_i has distance greater than 1 from u, then modify the subtree T' of T rooted at u as follows:

— add one node r_i to T' which becomes the new root of T' (and, hence, the new sibling of v_{i+1});

— make u the first child of r_i;

— let p be the parent of v_i; identify p with its first child and make v_i the second child of r_i.

It is easy to see that the resulting binary tree equals the syntax tree for the sequential compositional expression (12) which has exactly $2n-1$ nodes. Moreover, since each operation does not change the number of nodes of the current tree, we can conclude that the syntax tree for θ has exactly $2n-1$ nodes. □

6. EQUIVALENCE OF COMPOSITIONAL EXPRESSIONS

Let θ be a compositional expression with base sequence $\alpha_\theta = (X_1, X_2, \ldots, X_n)$. The system (i. e., the set) of sets $\mathcal{H}_\theta = \{X_1, X_2, \ldots, X_n\}$ will be referred to as the *base scheme* of θ. A set $\mathbf{d}$ of functions, one for each set X_i in $\mathcal{H}_\theta$, will be referred to as a *database* for θ if, for each set $X_i \in \mathcal{H}$, the function in $\mathbf{d}$ corresponding to X_i is a distribution function on X_i. Again, we assume that the distribution functions in $\mathbf{d}$ are all over the same semifield. Let $\mathbf{f} = (f_1, f_2, \ldots, f_n)$ be the ordering of the distribution functions in $\mathbf{d}$ according to α_θ, that is, f_i is a distribution function on X_i, $1 \le i \le n$; thus, $\mathbf{f}$ is an interpretation of θ, and we say that $\mathbf{d}$ is a *valid database* for θ if $\mathbf{f}$ is a valid interpretation of θ. Let E_θ be the operator that maps every valid database $\mathbf{d}$ for θ to the distribution function $[\theta]_\mathbf{f}$, where $\mathbf{f}$ is the (valid) interpretation of θ provided by $\mathbf{d}$, that is, $E_\theta(\mathbf{d}) = [\theta]_\mathbf{f}$. We call E_θ the *evaluation operator* of θ. Accordingly, the model M_θ generated by θ is the range of E_θ.

Two compositional expressions θ_1 and θ_2 with the same base scheme are said to be *equivalent* if they have the same evaluation operator, that is, if $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d})$ for every database $\mathbf{d}$ that is valid for both θ_1 and θ_2.

Example 6.1. Let $\mathbf{d} = \{f, g, h\}$ where f, g and h are distribution functions on AB, BCD and BCE, respectively. Consider the following three compositional expressions:

$$\theta_1 = AB \triangleright (BCD \triangleright BCE) \quad \theta_2 = (AB \triangleright BCD) \triangleright BCE \quad \theta_3 = (AB \triangleright BCE) \triangleright BCD.$$

Then, (f, g, h) is the interpretation provided by $\mathbf{d}$ for both θ_1 and θ_2, and (f, h, g) is the interpretation of θ_3 provided by $\mathbf{d}$. Assume that $\mathbf{d}$ is a valid database for all of them. Then, one has

$$E_{\theta_1}(\mathbf{d}) = f \times \frac{\frac{g \times h}{h^{\downarrow BC}}}{\sum_{C,D,E} \frac{g \times h}{h^{\downarrow BC}}} = f \times \frac{\frac{g \times h}{h^{\downarrow BC}}}{\sum_{C,D} g} = \frac{f \times g \times h}{g^{\downarrow B} \times h^{\downarrow BC}}.$$

Moreover, by eq. (17), one has

$$E_{\theta_2}(\mathbf{d}) = \frac{f \times g \times h}{g^{\downarrow B} \times h^{\downarrow BC}} \qquad E_{\theta_3}(\mathbf{d}) = \frac{f \times g \times h}{g^{\downarrow BC} \times h^{\downarrow B}}$$

Therefore, θ_1 is equivalent to θ_2 but is not equivalent to θ_3.

Note that Kratochvíl (see Remark 4.1 in [17]) calls "equivalent" two (sequential) compositional expressions θ_1 and θ_2 if $M_{\theta_1} = M_{\theta_2}$. The following simple example shows that our notion of equivalence is stronger than Kratochvíl's equivalence.

Example 6.2. Let $\mathbf{d} = \{f, g\}$ where f and g are distribution functions on X and Y, respectively. Consider the following two compositional expressions $\theta_1 = X \triangleright Y$ and $\theta_2 = Y \triangleright X$. Assume that $\mathbf{d}$ is a valid database for both θ_1 and θ_2. Then, one has

$$E_{\theta_1}(\mathbf{d}) = \frac{f \times g}{g^{\downarrow X \cap Y}} \qquad E_{\theta_2}(\mathbf{d}) = \frac{f \times g}{f^{\downarrow X \cap Y}}.$$

Since $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d})$ only for those databases $\mathbf{d}$ for which $f^{\downarrow X \cap Y} = g^{\downarrow X \cap Y}$, we can conclude that θ_1 and θ_2 are not equivalent. We now prove that $M_{\theta_1} = M_{\theta_2}$. To achieve this, it is sufficient to prove that there exists a valid database $\mathbf{d}'$ for θ_2 such that $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d}')$. Let $\mathbf{d}' = \{f, g'\}$ where g' is the distribution function on Y defined as follows:

$$g'(y) = \begin{cases} 0 & \text{if } g^{\downarrow X \cap Y}(y_{X \cap Y}) = 0 \\ g(y) \times \frac{f^{\downarrow X \cap Y}(y_{X \cap Y})}{g^{\downarrow X \cap Y}(y_{X \cap Y})} & \text{else.} \end{cases} \tag{14}$$

By the very definition of g' one has that if $g'(y) \neq 0$ then $f^{\downarrow X \cap Y}(y_{X \cap Y}) \neq 0$ which proves that $g' \propto f$ so that $\mathbf{d}'$ is a valid database for θ_2. At this point, it is easy to see that $E_{\theta_2}(\mathbf{d}') = g' \triangleright f = \frac{f \times g}{g^{\downarrow X \cap Y}} = E_{\theta_1}(\mathbf{d})$.

In the light of Example 6.2, a necessary condition for two compositional expressions to be equivalent is that their keys are the same.

Recall from the Introduction that a set system $\mathcal{H}$ is *star-like with centre* X if $X \in \mathcal{H}$ and, if $|\mathcal{H}| \geq 3$, then $Y \cap Z \subseteq X$ for every two distinct sets Y and Z in $\mathcal{H}$. If $\mathcal{H}$ is a star-like set system with centre X then, by Theorem 1.1, every two simple compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent. We shall prove a stronger result (see Theorem 6.5 below) which states that star-like set systems with centre X are precisely the set systems $\mathcal{H}$ for which every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent. To this end, we need the following lemma.

Lemma 6.3. Let $\mathcal{H}$ be a star-like set system with centre X, let θ be a compositional expression with base scheme $\mathcal{H}$ and key X, let $\alpha_\theta = (X_1 = X, X_2, \dots, X_n)$ and let $\mathbf{f} = (f_1, \dots, f_n)$ be a valid interpretation of θ. Then one has

$$[\theta]_{\mathbf{f}} = f_1 \times \prod_{2 \le i \le n} \frac{f_i}{f_i^{\downarrow X_1 \cap X_i}}. \tag{15}$$

Proof. First of all, let us consider the subexpression $\theta_0 = X_1 \triangleright (\theta_1)$ of θ. Thus $\alpha_{\theta_0} = (X_1 = X, X_2, \dots, X_k)$ and $\alpha_{\theta_1} = (X_2, \dots, X_k)$ for some k, $2 \le k \le n$. By Theorem 5.1, one has

$$[\theta_1]_{\mathbf{f}} = \frac{1}{q_{\theta_1}} \times \prod_{2 \le i \le k} f_i$$

where q_{θ_1} is either a function of COM(θ_1) or a constant, so that

$$[\theta_0]_{\mathbf{f}} = f_1 \times \frac{\prod_{2 \le i \le k} f_i}{q_{\theta_1} \times \sum_{A \in V(\theta_1) - X_1} \left(\frac{1}{q_{\theta_1}} \times \prod_{2 \le i \le k} f_i \right)}.$$

By hypothesis, $\mathcal{H}$ is star-like with centre X_1, so that

(*i*) COM$(\theta_1) \subseteq X_1$;

(*ii*) each variable in $V(\theta_1) - X_1$ occurs in exactly one X_i for some i, $2 \le i \le k$.

Since q_{θ_1} is either a function of COM(θ_1) or a constant, by (*i*) we can move $\frac{1}{q_{\theta_1}}$ to the left of the summation $\sum_{A \in V(\theta_1) - X_1}$:

$$[\theta_0]_{\mathbf{f}} = f_1 \times \frac{\prod_{2 \le i \le k} f_i}{q_{\theta_1} \times \frac{1}{q_{\theta_1}} \sum_{A \in V(\theta_1) - X_1} \prod_{2 \le i \le k} f_i}$$

$$= f_1 \times \frac{\prod_{2 \le i \le k} f_i}{\sum_{A \in V(\theta_1) - X_1} \prod_{2 \le i \le k} f_i}.$$

By (*ii*), one has

$$\sum_{A \in V(\theta_1) - X_1} \prod_{2 \le i \le k} f_i = \prod_{2 \le i \le k} f_i^{\downarrow X_1 \cap X_i}$$

so that

$$[\theta_0]_{\mathbf{f}} = f_1 \times \prod_{2 \le i \le k} \frac{f_i}{f_i^{\downarrow X_1 \cap X_i}}.$$

If $k = n$ then we are done. Otherwise, consider any subexpression θ' of θ containing θ_0 and let $\alpha_{\theta'} = (X_1, \dots, X_m)$, $k < m \le n$. By structural induction, it is easy to prove that

$$[\theta']_{\mathbf{f}} = f_1 \times \prod_{2 \le i \le m} \frac{f_i}{f_i^{\downarrow X_1 \cap X_i}}$$

and for $\theta' = \theta$ we obtain eq. (15). □

By Lemma 6.3 we have the following property of star-like set systems, which is a first generalization of Theorem 1.1.

Corollary 6.4. Let $\mathcal{H}$ be a star-like set system with centre X. Every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent.

Proof. Let $\mathcal{H} = \{X_1 = X, X_2, \ldots, X_n\}$ and let θ_1 and θ_2 be two compositional expressions with base scheme $\mathcal{H}$ and key X. Let $\alpha_{\theta_1} = (X_{h_1} = X, X_{h_2}, \ldots, X_{h_n})$ and $\alpha_{\theta_2} = (X_{k_1} = X, X_{k_2}, \ldots, X_{k_n})$ be the base sequences of θ_1 and θ_2, respectively. Let $\mathbf{d} = \{f_1, f_2, \ldots, f_n\}$ be a database in which f_i is a distribution function on X_i, $1 \leq i \leq n$. Assume that $\mathbf{d}$ is a valid database for both θ_1 and θ_2. By Lemma 6.3, one has

$$E_{\theta_1}(\mathbf{d}) = f_1 \times \prod_{2 \leq j \leq n} \frac{f_{h_j}}{f_{h_j}^{\downarrow X \cap X_{h_j}}} \qquad E_{\theta_2}(\mathbf{d}) = f_1 \times \prod_{2 \leq l \leq n} \frac{f_{k_l}}{f_{k_l}^{\downarrow X \cap X_{k_l}}}.$$

For each $i > 1$, let $j(i)$ and $l(i)$ be such that $X_i = X_{h_{j(i)}} = X_{k_{l(i)}}$. Then, for each $i > 1$ one has $f_{h_{j(i)}} = f_{k_{l(i)}} = f_i$ and $X \cap X_{h_{j(i)}} = X \cap X_{k_{l(i)}} = X \cap X_i$ so that $f_{h_{j(i)}}^{\downarrow X \cap X_{h_{j(i)}}} = f_{k_{l(i)}}^{\downarrow X \cap X_{k_{l(i)}}} = f_i^{\downarrow X \cap X_i}$. Therefore, $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d})$ and, hence, θ_1 and θ_2 are equivalent. □

We are now in a position to characterize the set systems $\mathcal{H}$ for which every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent.

Theorem 6.5. Let $\mathcal{H}$ be a set system, and let $X \in \mathcal{H}$. Every two compositional expressions with base scheme $\mathcal{H}$ and key X are equivalent if and only if $\mathcal{H}$ is star-like with centre X.

Proof. (*If*) By Corollary 6.4.
(*Only if*) Suppose, by contradiction, that $\mathcal{H}$ is not star-like with centre X. Then, $|\mathcal{H}| \geq 3$ and there exist two distinct sets Y and Z in $\mathcal{H} - \{X\}$ such that $Y \cap Z$ is not a subset of X. Let $\mathcal{H} = \{X_1, X_2, X_3, \ldots, X_n\}$, where $X_1 = X$, $X_2 = Y$ and $X_3 = Z$. We now prove that there exist two compositional expressions with base scheme $\mathcal{H}$ and key X_1 that are not equivalent so that a contradiction arises.

Without loss of generality, we assume that $n > 3$. (The case $n = 3$ can be proved using similar arguments.) Consider the following two sequential expressions with base scheme $\mathcal{H}$:

$$\theta_1 = (\ldots(((X_1 \triangleright X_2) \triangleright X_3) \triangleright X_4) \ldots) \triangleright X_n$$

$$\theta_2 = (\ldots(((X_1 \triangleright X_3) \triangleright X_2) \triangleright X_4) \ldots) \triangleright X_n.$$

We now prove that there exists a valid database $\bar{\mathbf{d}}$ for both θ_1 and θ_2 for which $E_{\theta_1}(\bar{\mathbf{d}}) \neq E_{\theta_2}(\bar{\mathbf{d}})$, which entails that θ_1 and θ_2 are not equivalent.

Let f_i denote any distribution function on X_i, $1 \leq i \leq n$; thus, f_1, f_2 and f_3 are distribution functions on $X_1 = X$, $X_2 = Y$ and $X_3 = Z$, respectively. Consider the set $\mathbf{D}$ of valid databases $\mathbf{d} = \{f_1, f_2, f_3, \ldots, f_n\}$ for both θ_1 and θ_2, in which f_3 is the unitary distribution function on Z. By eq. (13), for every $\mathbf{d} \in \mathbf{D}$ one has

$$E_{\theta_1}(\mathbf{d}) = f_1 \times \frac{f_2}{f_2^{\downarrow X_1 \cap X_2}} \times \frac{f_3}{f_3^{\downarrow (X_1 \cup X_2) \cap X_3}} \times \prod_{4 \le i \le n} \frac{f_i}{f_i^{\downarrow \partial_{\theta_1} X_i}}$$

$$E_{\theta_2}(\mathbf{d}) = f_1 \times \frac{f_2}{f_2^{\downarrow X_1 \cap X_3}} \times \frac{f_3}{f_3^{\downarrow (X_1 \cup X_3) \cap X_2}} \times \prod_{4 \le i \le n} \frac{f_i}{f_i^{\downarrow \partial_{\theta_2} X_i}}.$$

With

$$g = f_1 \times f_2 \times f_3 \times \prod_{4 \le i \le n} \frac{f_i}{f_i^{\downarrow \partial_{\theta_1} X_i}}$$

one has

$$E_{\theta_1}(\mathbf{d}) = \frac{g}{f_2^{\downarrow X \cap Y} \times f_3^{\downarrow (X \cup Y) \cap Z}}$$

and, since $\partial_{\theta_1} X_i = \partial_{\theta_2} X_i$ for each i, $4 \le i \le n$, one also has

$$E_{\theta_2}(\mathbf{d}) = \frac{g}{f_2^{\downarrow (X \cup Z) \cap Y} \times f_3^{\downarrow X \cap Z}}.$$

Moreover, since the distribution function f_3 is unitary, one has

$$f_3^{\downarrow X \cap Z} = |dom(Z - X)| \qquad f_3^{\downarrow (X \cup Y) \cap Z} = |dom(Z - (X \cup Y))|$$

so that

$$E_{\theta_1}(\mathbf{d}) = \frac{g}{f_2^{\downarrow X \cap Y} \times |dom(Z - (X \cup Y))|} \qquad E_{\theta_2}(\mathbf{d}) = \frac{g}{f_2^{\downarrow (X \cup Z) \cap Y} \times |dom(Z - X)|}.$$

Therefore, for every $\mathbf{d} \in \mathbf{D}$ one has $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d})$ if and only if f_2 is a solution of the following equation:

$$f_2^{\downarrow X \cap Y} \times |dom(Z - (X \cup Y))| = f_2^{\downarrow (X \cup Z) \cap Y} \times |dom(Z - X)|. \qquad (16)$$

Since, by hypothesis, $Y \cap Z$ is not a subset of X, one has

$$X \cap Y \neq (X \cup Z) \cap Y \qquad Z - (X \cup Y) \neq Z - X$$

so that eq. (16) is not an identity. Let

$$\bar{\mathbf{d}} = \{f_1, \bar{f}_2, f_3, f_4, \ldots, f_n\}$$

be any database in $\mathbf{D}$ in which $\bar{f}_2$ is not a solution of eq. (16). Then, $E_{\theta_1}(\bar{\mathbf{d}}) \neq E_{\theta_2}(\bar{\mathbf{d}})$, which proves that θ_1 and θ_2 are not equivalent (contradiction). □

7. CANONICAL EXPRESSIONS

A compositional expression θ with base sequence $\alpha_\theta = (X_1, \ldots, X_n)$ is a *canonical expression* if α_θ is a *perfect sequence* [18], that is, if the following property holds:

> (*running intersection property*) if $n > 1$ then, for each $i > 1$ there exists $j < i$ such that $(\cup_{1 \le h \le i-1} X_h) \cap X_i \subseteq X_j$.

Thus, using the notation introduced in Section 5, θ is a canonical expression if, for each $i > 1$, there exists $j < i$ such that $\partial_\theta X_i \subseteq X_j$.

Given a canonical expression θ with base sequence $\alpha_\theta = (X_1, \ldots, X_n)$, if $\mathbf{f} = (f_1, \ldots, f_n)$ is a valid interpretation of θ, then $[\theta]_\mathbf{f}$ has the following closed-form expression

$$[\theta]_\mathbf{f} = f_1 \times \prod_{2 \leq i \leq n} \frac{f_i}{f_i^{\downarrow \partial_\theta X_i}}. \tag{17}$$

The proof of eq. (17) was given in [22] and, for the sake of completeness, is reported in the Appendix. Note that eq. (17) is the same as eq. (13) and does mean that $[\theta]_\mathbf{f}$ depends on θ only through the base sequence α_θ of θ, more precisely, through the sets $X_1, \partial_\theta X_2, \ldots, \partial_\theta X_n$.

We shall state a necessary and sufficient condition for the equivalence of two canonical expressions with the same base scheme and the same key. Next, we shall characterize those set systems $\mathcal{H}$ for which every two canonical expressions with base scheme $\mathcal{H}$ and with the same key, say X, are equivalent. We call such set systems "X-centric", and prove that the class of X-centric set systems strictly includes the class of star-like set systems with centre X. First of all, we recall some useful notions related to canonical expressions.

7.1. Acyclic hypergraphs

A *hypergraph* is a system (that is, a set) of distinct nonempty sets. If $\mathcal{H}$ is a hypergraph, by $V(\mathcal{H})$ we denote the union of the sets in $\mathcal{H}$. A hypergraph $\mathcal{H}$ is *acyclic* if there exists an ordering of the sets in $\mathcal{H}$ which is a perfect sequence (see Section 7). Such orderings of an acyclic hypergraph $\mathcal{H}$ are called *perfect orderings* of $\mathcal{H}$.
Of course, the base scheme of a canonical expression is always an acyclic hypergraph, but a compositional expression whose base scheme is an acyclic hypergraph need not be a canonical expression. Moreover, if $\mathcal{H}$ is a star-like set system with centre X, then every ordering $(X_1, X_2, \ldots, X_n)$ of $\mathcal{H}$ with $X_1 = X$ (or with $X_2 = X$) is perfect, which proves that $\mathcal{H}$ is an acyclic hypergraph.

Acyclic hypergraphs are also called *decomposable hypergraphs* [18] and *hypertrees* [28], and an efficient algorithm to test acyclicity of hypergraphs can be found in [29]. There exist several characterizations of acyclic hypergraphs exist [1, 18]. We now recall one of them, which serves our purpose.

A *junction tree* [18] (or *join tree* [1] or *clique tree* [2, 6] or *Markov tree* [28]) of a hypergraph $\mathcal{H}$ is an undirected tree J with node set $\mathcal{H}$ in which, for every two distinct nodes X and Y

> (*junction property*) for every edge (E, F) is that is along the X-Y path (that is, along the unique path joining X and Y in J), one has $X \cap Y \subseteq E \cap F$.

Theorem 7.1. (Beeri et al. [1], Lauritzen [18]) A hypergraph is acyclic if and only if it has a junction tree.

An efficient algorithm for constructing a junction tree of an acyclic hypergraph can be found in [29]. We now recall a useful property of junction trees.

Let $\mathcal{H}$ be any hypergraph. Let X and Y be two distinct sets in $\mathcal{H}$. An X-Y *chain* in $\mathcal{H}$ is a sequence $(E_1, \ldots, E_n)$ of distinct sets in $\mathcal{H}$ such that $E_1 = X$, $E_n = Y$ and $E_i \cap E_{i+1} \neq \emptyset$, for each $i < n$. Two sets X and Y in $\mathcal{H}$ are *connected* if either $X = Y$ or there exists an X-Y chain in $\mathcal{H}$. A subset S of $V(\mathcal{H})$ such that $X - S \neq \emptyset$ and $Y - S \neq \emptyset$, is an X-Y *separator* of $\mathcal{H}$ if either X and Y in $\mathcal{H}$ are not connected or, for every X-Y chain $(E_1, \ldots, E_n)$, there exists i for which $E_i \cap E_{i+1} \subseteq S$.

Consider now an acyclic hypergraph $\mathcal{H}$, and let J be a junction tree of $\mathcal{H}$. A path p in J corresponds to a chain in $\mathcal{H}$ if and only if there is no edge (E, F) along p for which $E \cap F = \emptyset$. In other words, two distinct sets X and Y in $\mathcal{H}$ are connected if and only if there is a junction tree J of $\mathcal{H}$ in which, for every edge (E, F) along the X-Y path, one has $E \cap F \neq \emptyset$. The following is a well-known (e. g., see [2]) characterization of separators of an acyclic hypergraph.

Lemma 7.2. Let $\mathcal{H}$ be an acyclic hypergraph, let X and Y, $X \neq Y$, be two distinct sets in $\mathcal{H}$, and let J be a junction tree of $\mathcal{H}$. A subset S of $V(\mathcal{H})$ is an X-Y separator of $\mathcal{H}$ if and only if $X - S \neq \emptyset$, $Y - S \neq \emptyset$ and there is an edge (E, F) along the X-Y path in J such that $E \cap F \subseteq S$.

Let $\mathcal{H}$ be an acyclic hypergraph, let J be a junction tree of $\mathcal{H}$, and let X be any node of J. An X-*rooted junction tree* of $\mathcal{H}$ is the directed tree T obtained by rooting J at the node X and orienting the edges of J away from the root X. Thus, an edge (E, F) is ordiented from E to F, written $E \to F$, if in J the distance of E from X is less than the distance of F from X. To avoid ambiguity, we call the oriented edge $E \to F$ an *arc* of T; moreover, we say that E is the *parent* of F in T, written $E = pa_T(F)$. Let Y and Z be two distinct nodes of T; by $LCA(Y, Z)$ we denote the lowest common ancestor of Y and Z in T. Again, T enjoys the junction property which now reads: For every two distinct nodes Y and Z of T,

— if $Y = LCA(Y, Z)$ (or $Z = LCA(Y, Z)$) then, for every arc $E \to F$ of T that is along the directed path from Y to Z (from Z to Y, respectively), one has $Y \cap Z \subseteq E \cap F$;

— if $LCA(Y, Z) \notin \{Y, Z\}$ then, for every arc $E \to F$ of T that is along the directed path from $LCA(Y, Z)$ to Y or along the directed path from $LCA(Y, Z)$ to Z, one has $Y \cap Z \subseteq E \cap F$.

There exists a many-to-many correspondence between X-rooted junction trees of an acyclic hypergraph $\mathcal{H}$ and perfect orderings of $\mathcal{H}$ beginning with X.

Given an X-rooted junction tree T of $\mathcal{H}$, top-down (i. e., root-to-leaf) traversals of T generate perfect orderings of $\mathcal{H}$. If α is such a perfect ordering of $\mathcal{H}$ then, for every non-root node Y of T, one has $\partial_\alpha Y = Y \cap pa_T(Y)$, and we call α a *perfect ordering* of $\mathcal{H}$ *associated* with T. On the other hand, given a perfect ordering $\alpha = (X_1 = X, \ldots, X_n)$ of $\mathcal{H}$, an X-rooted junction tree of $\mathcal{H}$ can be obtained as follows. For each $i > 1$, take $pa_T(X_i)$ to be one of the nodes X_j, $j < i$, for which $\partial_\alpha X_i \subseteq X_j$. We call a tree such as T an X-*rooted junction tree* of $\mathcal{H}$ *associated* with α.

Corollary 7.3. Let $\mathcal{H}$ be an acyclic hypergraph, and let X and Y, $X \neq Y$, be two sets in $\mathcal{H}$. Let T_1 and T_2 be two X-rooted junction trees of $\mathcal{H}$, and let $S_h = Y \cap pa_{T_h}(Y)$,

$h = 1, 2$. There is an arc $E \to F$ along the directed path from X to Y in T_1 such that $E \cap F \subseteq S_1 \cap S_2$.

Proof. Let p_h be the directed path from X to Y in T_h, $h = 1, 2$. The statement is obvious if X and Y are not connected in $\mathcal{H}$ for, then, there is an arc $E \to F$ along p_1 such that $E \cap F = \emptyset$. Assume that X and Y are connected in $\mathcal{H}$. Let us distinguish the following three cases.

Case 1: $X \subseteq S_2$. In this case, $X \subset Y$ and, by the junction property of T_1, each node along p_1 is a set that contains X; therefore, X is a subset of $pa_{T_1}(Y)$ and, hence, of $Y \cap pa_{T_1}(Y) = S_h$. So, $X \subseteq S_1 \cap S_2$. Let $E = X$ and let F be the child of X on p_1; then, the arc $E \to F$ is such that $E \cap F = X \subseteq S_1 \cap S_2$.

Case 2: $S_2 = Y$. In this case, since $S_1 \subseteq Y$, one has $S_1 \cap S_2 = S_1$. Let $E = pa_{T_1}(Y)$ and $F = Y$; then, the arc $E \to F$ is such that $E \cap F = S_1 = S_1 \cap S_2$.

Case 3: $X - S_2 \neq \emptyset$ and $Y - S_2 \neq \emptyset$. Since the arc $pa_{T_2}(Y) \to Y$ of T_2 is along p_2 and $S_2 = Y \cap pa_{T_2}(Y)$, by Lemma 7.2 applied to T_2 the set S_2 is an X-Y separator of $\mathcal{H}$. Therefore, by Lemma 7.2 applied to T_1, there is an arc $E \to F$ along p_1 such that $E \cap F \subseteq S_2$. On the other hand, since $S_2 \subset Y$, one has that $E \cap F \subset Y$ and, hence $E \cap F \subseteq E \cap Y$. By the junction property of T_1, $E \cap Y \subseteq pa_{T_1}(Y) \cap Y = S_1$ so that $E \cap F \subseteq E \cap Y \subseteq S_1$. To sum up, the arc $E \to F$ of T_1 is such that $E \cap F \subseteq S_1 \cap S_2$. □

Example 7.4. Consider the acyclic hypergraph $\mathcal{H} = \{AF, ABC, ABE, ACD\}$, and let $X = AF$ and $Y = ABC$. Let T_1 be the X-rooted junction tree of $\mathcal{H}$ in which the directed path from X to Y is $p_1 = (AF, ABE, ABC)$, and let T_2 be the X-rooted junction tree of $\mathcal{H}$ in which the directed path from X to Y is $p_2 = (AF, ACD, ABC)$. Then $pa_{T_1}(Y) = ABE$ and $pa_{T_2}(Y) = ACD$, so that $S_1 = Y \cap pa_{T_1}(Y) = AB$, $S_2 = Y \cap pa_{T_2}(Y) = AC$ and $S_1 \cap S_2 = A$. The arc $AF \to ABE$ along p_1 is such that $AF \cap ABE \subseteq S_1 \cap S_2$, and the arc $AF \to ACD$ along p_2 is such that $AF \cap ACD \subseteq S_1 \cap S_2$.

7.2. An equivalence criterion

The proof of the following result is similar to the proof of Theorem 6.5.

Theorem 7.5. Let $\mathcal{H}$ be an acyclic hypergraph, and let $X \in \mathcal{H}$. Two canonical expressions θ_1 and θ_2 with base scheme $\mathcal{H}$ and key X are equivalent if and only if, for every $Y \in \mathcal{H} - \{X\}$, one has $\partial_{\theta_1} Y = \partial_{\theta_2} Y$.

Proof. (*If*) By eq. (17), θ_1 and θ_2 are equivalent.
(*Only if*) By hypothesis, θ_1 and θ_2 are equivalent. Suppose, by contradiction, that there exists $Y \in \mathcal{H} - \{X\}$ such that $\partial_{\theta_1} Y \neq \partial_{\theta_2} Y$. We now prove that there exists a valid database $\bar{\mathbf{d}}$ for both θ_1 and θ_2 such that $E_{\theta_1}(\bar{\mathbf{d}}) \neq E_{\theta_2}(\bar{\mathbf{d}})$ so that a contradiction arises. Let $\mathcal{H} = \{X_1 = X, X_2 = Y, X_3, \ldots, X_n\}$, $n \geq 3$. Consider the set $\mathbf{D}$ of valid databases $\mathbf{d} = \{f_1, f_2, f_3, \ldots, f_n\}$ for both θ_1 and θ_2 in which, for each $i \neq 2$, f_i is a unitary distribution function on X_i. For each $i > 2$, since f_i is a unitary distribution function, one has

$$f_i^{\downarrow \partial_{\theta_h} X_i} = |dom(X_i - \partial_{\theta_h} X_i)| \qquad (h = 1, 2).$$

Therefore, by eq. (17), for every $\mathbf{d} \in \mathbf{D}$ one has

$$E_{\theta_h}(\mathbf{d}) = \frac{1}{m_h} \times \frac{f_2}{f_2^{\downarrow \partial_{\theta_h} Y}} \qquad (h = 1, 2)$$

where

$$m_h = \prod_{3 \leq i \leq n} |dom(X_i - \partial_{\theta_h} X_i)| \qquad (h = 1, 2).$$

So, for every $\mathbf{d} \in \mathbf{D}$, $E_{\theta_1}(\mathbf{d}) = E_{\theta_2}(\mathbf{d})$ if and only if f_2 is a solution of the following equation:

$$m_1 \times f_2^{\downarrow \partial_{\theta_1} Y} = m_2 \times f_2^{\downarrow \partial_{\theta_2} Y}$$

which is not an identity since $\partial_{\theta_1} Y \neq \partial_{\theta_2} Y$. Let $\bar{\mathbf{d}} = \{f_1, \bar{f}_2, f_3, \ldots, f_n\}$ be a database in $\mathbf{D}$ in which $\bar{f}_2$ is not a solution of the equation above. Then, $E_{\theta_1}(\bar{\mathbf{d}}) \neq E_{\theta_2}(\bar{\mathbf{d}})$ and, hence, a contradiction arises. □

Theorem 7.5 can be re-phrased in graphical terms using tree representations of perfect sequences mentioned in Subsection 7.1.

Corollary 7.6. Let $\mathcal{H}$ be an acyclic hypergraph, and let $X \in \mathcal{H}$. Let θ_1 and θ_2 be two canonical expressions with base scheme $\mathcal{H}$ and key X, and let T_h be an X-rooted junction tree of $\mathcal{H}$ associated with the base sequence of θ_h, $h = 1, 2$. The expressions θ_1 and θ_2 are equivalent if and only if, for every $Y \in \mathcal{H} - \{X\}$, one has $Y \cap pa_{T_1}(Y) = Y \cap pa_{T_2}(Y)$.

P r o o f. By Theorem 7.5, θ_1 and θ_2 are equivalent if and only if, for every $Y \in \mathcal{H} - \{X\}$, one has $\partial_{\theta_1} Y = \partial_{\theta_2} Y$. The statement then follows from the hypothesis that T_h is an X-rooted junction tree of $\mathcal{H}$ associated with the base sequence of θ_h, which implies that $\partial_{\theta_h} Y = Y \cap pa_{T_h}(Y)$, $h = 1, 2$. □

7.3. X-centric set systems

Let $\mathcal{H}$ be an acyclic hypergraph and let $X \in \mathcal{H}$. We say that $\mathcal{H}$ is an *X-centric set system* if, for every two X-rooted junction trees T_1 and T_2 of $\mathcal{H}$, one has that $Y \cap pa_{T_1}(Y) = Y \cap pa_{T_2}(Y)$ for every $Y \in \mathcal{H} - \{X\}$. Note that, if $\mathcal{H}$ is star-like with centre X, then $\mathcal{H}$ is an X-centric set system.

Example 7.7. The acyclic hypergraph $\mathcal{H} = \{AB, BCD, BCE\}$ has two AB-rooted junction trees T_1 and T_2: the arcs of T_1 are $AB \to BCD$ and $BCD \to BCE$, and the arcs of T_2 are $AB \to BCE$ and $BCE \to BCD$. Since $BCD \cap pa_{T_1}(BCD) \neq BCD \cap pa_{T_2}(BCD)$, $\mathcal{H}$ is not an AB-centric set system.

Theorem 7.8. Let $\mathcal{H}$ be an acyclic hypergraph and let $X \in \mathcal{H}$. Every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent if and only if $\mathcal{H}$ is an X-centric set system.

P r o o f. (*If*) Assume that $\mathcal{H}$ is an X-centric set system. Let θ_1 and θ_2 be any two canonical expressions with base scheme $\mathcal{H}$ and key X, and let T_h be an X-rooted junction tree of $\mathcal{H}$ associated with the base sequence of θ_h, $h = 1, 2$. Since $\mathcal{H}$ is an X-centric hypergraph, for each $Y \in \mathcal{H} - \{X\}$ one has that $Y \cap pa_{T_1}(Y) = Y \cap pa_{T_2}(Y)$. By Corollary 7.6, θ_1 and θ_2 are equivalent.
(*Only if*) Assume that every two canonical expressions with base scheme $\mathcal{H}$ and key X are equivalent and suppose, by contradiction, that $\mathcal{H}$ is not an X-centric set system. Then, there exist two X-rooted junction trees T_1 and T_2 of $\mathcal{H}$ and a set $Y \in \mathcal{H} - \{X\}$ such that $Y \cap pa_{T_1}(Y) \neq Y \cap pa_{T_2}(Y)$. Let α_h be a perfect ordering of $\mathcal{H}$ associated with T_h, and let θ_h be a canonical expression with base scheme $\mathcal{H}$ and base sequence α_h, $h = 1, 2$. By Corollary 7.6, θ_1 and θ_2 are not equivalent (contradiction). □

The next theorem provides an efficient algorithm to recognize X-centric set system.

Theorem 7.9. Let $\mathcal{H}$ be an acyclic hypergraph, let $X \in \mathcal{H}$ and let T be an X-rooted junction tree of $\mathcal{H}$. $\mathcal{H}$ is an X-centric set system if and only if T enjoys the following property

(π) For every interior node $Y \neq X$ of T, there is no child Z of Y such that $Y \cap pa_T(Y) \subset Y \cap Z$ (that is, $Y \cap pa_T(Y)$ is a proper subset of $Y \cap Z$).

P r o o f. (*Only if*) Assume that $\mathcal{H}$ is an X-centric set system and suppose, by contradiction, that there exist an interior node $Y \neq X$ of T and a child Z of Y such that $Y \cap pa_T(Y)$ is a proper subset of $Y \cap Z$. Let $P = pa_T(Y)$. Since $Y \cap P \subset Y \cap Z \subseteq Z$, one has $Y \cap P \subseteq Z \cap P$. On the other hand, by the junction property of T, one has that $Z \cap P \subseteq Y$ and, hence, $Z \cap P \subseteq Y \cap P$. It follows that $Y \cap P = Z \cap P$. At this point, we can construct an X-rooted tree T' of $\mathcal{H}$ from T by replacing the arcs $P \to Y$ and $Y \to Z$ by $P \to Z$ and $Z \to Y$. It is easy to see that T' is an X-rooted junction tree of $\mathcal{H}$. Then, since $Y \cap P \subset Y \cap Z$ by hypothesis, one has $Y \cap P \neq Y \cap Z$, that is, $Y \cap pa_T(Y) \neq Y \cap pa_{T'}(Y)$ which proves that $\mathcal{H}$ is not an X-centric set system (contradiction).
(*If*) Assume that T enjoys property (π) and suppose, by contradiction, that there exists another X-rooted junction tree T' of $\mathcal{H}$ containing a node $Y \neq X$ such that $Y \cap pa_T(Y) \neq Y \cap pa_{T'}(Y)$. Let $P = pa_T(Y)$ and $P' = pa_{T'}(Y)$, and let $S = Y \cap P$ and $S' = Y \cap P'$. So, $S \neq S'$. Let us distinguish two cases depending on whether $S \subset S'$ or $S - S' \neq \emptyset$.

Case 1: $S \subset S'$. In T the node P' must be a descendant of Y for, otherwise, in the junction tree J underlying T P should be on the Y-P' path and, by the junction property of J, $Y \cap P'$ should be a subset of P and, hence, one would have $S' \subseteq S$ and, since $S \neq S'$, $S' \subset S$ (contradiction). Let Z be the child of Y that lies on the directed path in T from Y to P'. By the junction property of T, one has that $Y \cap P' \subseteq Y \cap Z$ and, since $S \subset S'$, one has that $S \subset S' = Y \cap P' \subset Y \cap Z$ and, hence, the node Y violates condition (π) (contradiction).

Case 2: $S - S' \neq \emptyset$. First of all, observe that in this case $S \cap S'$ is a proper subset of S. Let p be the directed path in T from X to Y. By Corollary 7.3, there exists an arc $E \to F$ along p such that $E \cap F \subseteq S \cap S'$. Let $E^* \to F^*$ be the deepest (that is, the nearest to Y) of such arcs, and let $L = E^* \cap F^*$; so, $L \subseteq S \cap S'$ and, since $S \cap S' \subset S$, one has $L \subset S$ so that $E^* \to F^* \neq P \to Y$. Of course, L is a subset of F and, since $S \subseteq Y$, L is also a subset of Y so that $L \subseteq F^* \cap Y$. Let Z be the child of F^* along p. By the junction property of T, $F^* \cap Y \subseteq F^* \cap Z$ and, hence, $L \subseteq F^* \cap Z$. The equality cannot hold for, otherwise, $F^* \cap Z = L \subseteq S \cap S'$ and $E^* \to F^*$ wouldn't be the deepest of the arcs $E \to F$ along p for which $E \cap F \subseteq S \cap S'$. So, $L \subset F^* \cap Z$ which proves that the node F^* violates condition (π) (contradiction). □

8. DETECTION OF CONDITIONAL INDEPENDENCES

Let θ be a compositional expression. A conditional independence $X \perp\!\!\!\perp Y \mid Z$ *holds in* the model M_θ generated by θ if it holds in the distribution function $[\theta]_{\mathbf{f}}$, for every valid interpretation $\mathbf{f}$ of θ. Which conditional independences hold in M_θ?

Let X be the key of θ, and let $(\theta_1, \ldots, \theta_k)$ be the sequence of distinct subexpressions of θ such that each θ_i contains X and for each i, $1 \leq i \leq k-1$, θ_i is a subexpression of θ_{i+1}. Thus, $\theta_1 = X$ and $\theta_k = \theta$. Let $\mathbf{f}$ be a valid interpretation of θ. By Fact 4.3, for each $i < k$, $[\theta_i]_{\mathbf{f}}$ is the marginal of $[\theta_{i+1}]_{\mathbf{f}}$ on $V(\theta_i)$ and, since $[\theta_k]_{\mathbf{f}} = [\theta]_{\mathbf{f}}$, $[\theta_i]_{\mathbf{f}}$ is the marginal of $[\theta]_{\mathbf{f}}$ on $V(\theta_i)$. Moreover, by Theorem 4.4, for each $i < k$, if $V(\theta_i) - V(\theta_{i+1}) \neq \emptyset$ and $V(\theta_{i+1}) - V(\theta_i) \neq \emptyset$ then the conditional independence

$$V(\theta_i) - V(\theta_{i+1}) \quad \perp\!\!\!\perp \quad V(\theta_{i+1}) - V(\theta_i) \quad \mid \quad V(\theta_i) \cap V(\theta_{i+1}) \tag{18}$$

holds in $[\theta_{i+1}]_{\mathbf{f}}$ and, hence, in the marginal of $[\theta]_{\mathbf{f}}$ on $V(\theta_{i+1})$. So, each conditional independence (18) holds in $[\theta]_{\mathbf{f}}$ and, hence, in M_θ.

Consider now the case that θ is a sequential compositional expression with base sequence $\alpha_\theta = (X_1, \ldots, X_n)$, $n > 1$. Then, $\theta_i = (\ldots(X_1 \rhd X_2) \rhd \ldots) \rhd X_i$ for each $i > 1$, each conditional independence (18) reduces to

$$V(\theta_i) - \partial_\theta X_{i+1} \quad \perp\!\!\!\perp \quad X_{i+1} - \partial_\theta X_{i+1} \quad \mid \quad \partial_\theta X_{i+1}. \tag{19}$$

Moreover, by the decomposition axiom, one has that, for each $A \in V(\theta_i) - \partial_\theta X_{i+1}$ and each $B \in X_{i+1} - \partial_\theta X_{i+1}$, the conditional independence

$$A \perp\!\!\!\perp B \mid \partial_\theta X_{i+1}$$

holds in M_θ.

Finally, consider the case that θ is a canonical expression with base sequence $\alpha_\theta = (X_1, \ldots, X_n)$, $n > 1$; thus, α_θ is a perfect sequence. Let θ' be the sequential compositional expression with $\alpha_{\theta'} = \alpha_\theta$, and let $\mathbf{f}$ be a valid interpretation of both θ and θ'. Since $\alpha_{\theta'} = \alpha_\theta$, by Theorem 10.5 one has $[\theta]_{\mathbf{f}} = [\theta']_{\mathbf{f}}$, so that all the conditional independences (19) hold in M_θ. As an application, we can answer the following question [10]: given two variables A and B in $V(\theta)$, which are the minimal (with respect to set-inclusion) sets S (if any) such that the conditional independence $A \perp\!\!\!\perp B \mid S$ holds in M_θ? The solution algorithm is as follows. Given a junction tree J of $\mathcal{H}_\theta$, find the shortest path $(E_1, \ldots, E_k)$ in T such that $A \in E_1$ and $B \in E_k$. If $k = 1$ (that is, if both A and B belong to E_1) then there exists no subset S of $V(\theta) - AB$ for which $A \perp\!\!\!\perp B \mid S$ holds in M_θ. Otherwise (that is, if $k > 1$), the minimal sets in $\{E_h \cap E_{h+1} : 1 \leq h \leq k-1\}$

are precisely the minimal sets S for which the conditional independence $A \perp\!\!\!\perp B \mid S$ holds in M_θ.

Before closing this section, we mention that, for a generating sequence of probability distributions, an effective method for detecting conditional independences was given by Jiroušek [9, 10] based on a tabular representation called *persegram*. Structural properties of persegrams are stated by Kratochvíl [16], who in [17] provides a method to decide, given two generating sequences of probability distributions, whether or not the conditional independences in the two compositional models are the same.

9. CLOSING NOTE

We have presented an extension of compositional model theory in two ways. First, we consider general distribution functions, whose class includes not only probability distributions but more in general multivariate functions whose values can be added, multiplied and divided. Second, we consider models generated by compositional expressions, whose class includes both simple and sequential compositional expressions. In order to assess the power of our compositional-expression formalism, we need to answer the following questions. How many are the sequential compositional expressions with a given base scheme? and the simple compositional expressions? and the compositional expressions? We now answer these questions for compositional expressions having in common a base scheme $\mathcal{H}$ with n sets.

It is easy to see that the number of sequential compositional expressions with base scheme $\mathcal{H}$ is $n!$

Consider now simple compositional expressions with base scheme $\mathcal{H}$. First of all, observe that, for a fixed ordered couple (X, Y) of distinct sets in $\mathcal{H}$, for each i, $1 \leq i \leq n-1$, there exist exactly $(n-2)!$ simple compositional expressions θ that contain the subexpression $(X \rhd Y)$ and in which X is the ith term of α_θ. Therefore, since the number of ordered couples such as (X, Y) is $n \cdot (n-1)$, the total number of simple compositional expressions with base scheme $\mathcal{H}$ is

$$n \cdot (n-1) \cdot \big((n-1) \cdot (n-2)!\big) = (n-1) \cdot n!$$

Finally, consider general compositional expressions with base scheme $\mathcal{H}$. We can determine the number e_n of such compositional expressions by means of a recurrence relation of the type $e_{n+1} = f(e_n)$ with $e_2 = 2$. In order to obtain such a recurrence relation, consider compositional expressions with base scheme $\mathcal{H} \cup \{X\}$ for $X \notin \mathcal{H}$. Each of them can be obtained from a compositional expression θ with base scheme $\mathcal{H}$ by replacing any subexpression θ' of θ by either $(\theta') \rhd X$ or $X \rhd (\theta')$. Since the number of subexpressions of θ is $2n-1$ by Theorem 5.4, we generate $2 \cdot (2n-1)$ distinct compositional expressions with base scheme $\mathcal{H} \cup \{X\}$ from each compositional expression with base scheme $\mathcal{H}$; moreover, the compositional expressions with base scheme $\mathcal{H} \cup \{X\}$ generated from two distinct compositional expressions with base scheme $\mathcal{H}$ are distinct too. Therefore, we can write down the following recurrence relation:

$$e_{n+1} = 2 \cdot (2n-1) \cdot e_n$$

so that

$$\begin{aligned} e_{n+1} &= 2 \cdot (2n-1) \cdot e_n = 2^2 \cdot (2n-1) \cdot (2n-3) \cdot e_{n-1} \\ &= 2^3 \cdot (2n-1) \cdot (2n-3) \cdot (2n-5) \cdot e_{n-2} = \dots \\ &= 2^{k+1} \cdot (2n-1) \cdot (2n-3) \cdot \ldots \cdot (2n-2k-1) \cdot e_{n-k}. \end{aligned}$$

For $n-k=2$, we know that $e_2=2$ so that for $k=n-2$ we obtain

$$e_{n+1} = 2^{n-1} \cdot (2n-1) \cdot (2n-3) \cdot \ldots \cdot 3 \cdot 2$$

$$= 2^n \cdot (2n-1) \cdot (2n-3) \cdot \ldots \cdot 3 = 2^n \cdot (2n-1)!! = (n+1)! \cdot C_n$$

where $C_n = 2^n \cdot \frac{(2n-1)!!}{(n+1)!}$ is the nth Catalan number. So, for each $n > 1$, one has

$$e_n = n! \cdot C_{n-1} = \frac{2 \cdot (2n-3)!}{(n-2)!}.$$

The following table reports the number s_n of sequential compositional expressions, the number s_n^* of simple compositional expressions, and the number e_n of compositional expressions, for $n = 2,3,4,5$.

n	s_n	s_n^*	e_n
2	2	2	2
3	6	12	12
4	24	72	120
5	120	480	1680

Before closing this section, we want to mention a number of open problems left to future research:

— the problem of testing the equivalence of any two compositional expressions with the same base scheme and the same key;

— the recognition of "multiplicative models", by which we mean compositional models for which there exists a closed-form formula (e. g., models generated by sequential compositional expressions or by canonical expressions);

— efficient procedures for marginalization in compositional models;

— a general form of composition expression in which a set can appear more than once.

10. APPENDIX

In order to prove eq. (17), we first provide a formula for $[\theta']_{\mathbf{f}}$ (see eq. (21) below) for any subexpression θ' of θ. To achieve this, we need some useful notations and two technical lemmas. Let $\alpha_{\theta'} = (X_k, X_{k+1}, \dots, X_m)$ be the base sequence of θ', for some k and m, $1 \le k \le m \le n$. It is convenient to partition the base scheme $\mathcal{H}_{\theta'} = \{X_k, X_{k+1}, \dots, X_m\}$ of θ' into two subsystems $\mathcal{R}_{\theta'}$ and $\mathcal{S}_{\theta'}$ which are defined as follows:

$$\mathcal{R}_{\theta'} = \{X_i \in \mathcal{H}_{\theta'} : \exists\, X_j \in \mathcal{H}_{\theta'}, j < i,\ X_j \cap X_i = \partial_\theta X_i\}$$
$$\mathcal{S}_{\theta'} = \mathcal{H}_{\theta'} - \mathcal{R}_{\theta'}\ . \tag{20}$$

Note that one always has $\mathcal{S}_{\theta'} \neq \emptyset$ since $X_k \in \mathcal{S}_{\theta'}$. By $V(\mathcal{S}_{\theta'})$ we denote the union of the sets in $\mathcal{S}_{\theta'}$, by COM($\mathcal{S}_{\theta'}$) the set of variables that are common to at least two distinct sets in $\mathcal{S}_{\theta'}$, and by UNI($\mathcal{S}_{\theta'}$) the complement of COM($\mathcal{S}_{\theta'}$) in $V(\mathcal{S}_{\theta'})$.

Example 10.1. Consider the canonical expression

$$\theta = ABCD \rhd (((ABE \rhd BCF) \rhd FL) \rhd ((CDG \rhd ADH) \rhd (CI \rhd BCM))).$$

Then

$$\alpha_\theta = (ABCD, ABE, BCF, FL, CDG, ADH, CI, BCM)$$

and

ABE	$\partial_\theta ABE = AB$
BCF	$\partial_\theta BCF = BC$
FL	$\partial_\theta FL = F$
CDG	$\partial_\theta CDG = CD$
ADH	$\partial_\theta ADH = AD$
CI	$\partial_\theta CI = C$
BCM	$\partial_\theta BCM = BC$

For the following three subexpressions of θ:

$$\theta_1 = (ABE \rhd BCF) \rhd FL$$

$$\theta_2 = (CDG \rhd ADH) \rhd (CI \rhd BCM)$$

$$\theta_3 = (\theta_1) \rhd (\theta_2)$$

we have

$$\mathcal{H}_{\theta_1} = \{ABE, BCF, FL\}$$

$$\mathcal{H}_{\theta_2} = \{CDG, ADH, CI, BCM\}$$

$$\mathcal{H}_{\theta_3} = \{ABE, BCF, FL, CDG, ADH, CI, BCM\}$$

and

h	$\mathcal{R}_{\theta_h}$	$\mathcal{S}_{\theta_h}$	COM($\mathcal{S}_{\theta_h}$)
1	$\{FL\}$	$\{ABE, BCF\}$	B
2	$\{CI\}$	$\{CDG, ADH, BCM\}$	CD
3	$\{FL, CI, BCM\}$	$\{ABE, BCF, CDG, ADH\}$	$ABCD$

The next lemma states useful properties of $\mathcal{R}_{\theta'}$ and $\mathcal{S}_{\theta'}$ for any subexpression θ' of a canonical expression.

Lemma 10.2. Let θ be a canonical expression, and let $\alpha_\theta = (X_1, \ldots, X_n)$. Let θ' be a subexpression of θ, let $\alpha_{\theta'} = (X_k, \ldots, X_m)$, and let $(\mathcal{R}_{\theta'}, \mathcal{S}_{\theta'})$ be the bipartition of $\mathcal{H}_{\theta'}$ defined by eq. (20).

(a) For each $X_i \in \mathcal{R}_{\theta'}$, one has that $(X_i - \partial_\theta X_i) \cap V(\mathcal{S}_{\theta'}) = \emptyset$.

(b) For each $X_i \in \mathcal{S}_{\theta'} - \{X_1\}$, $X_i - \partial_\theta X_i \subseteq \mathrm{UNI}(\mathcal{S}_{\theta'})$.

Proof. (a) Let X_i be any set in $\mathcal{R}_{\theta'}$. Suppose, by contradiction, that there exists a variable A in $X_i - \partial_\theta X_i$ that also belongs to some $X_j \in \mathcal{S}_{\theta'}$. Since $\partial_\theta X_i = (\cup_{1 \le l < i} X_l) \cap X_i$, one has that $j > i$ so that $A \in X_i \cap X_j \subseteq \partial_\theta X_j$. Moreover, since $X_j \in \mathcal{S}_{\theta'}$ and α_θ enjoys the running intersection property, there exists $h < k$ for which $\partial_\theta X_j = X_h \cap X_j$ so that $A \in X_h$; but, since $h < k \le i$ one has $A \in X_h \cap X_i$ and, hence, $A \in \partial_\theta X_i$ (contradiction).
(b) Let X_i be any set in $\mathcal{S}_{\theta'} - \{X_1\}$. Suppose, by contradiction, that there exists a variable A in $X_i - \partial_\theta X_i$ that also belongs to some $X_j \in \mathcal{S}_{\theta'}$, $j \neq i$. Since $\partial_\theta X_i = (\cup_{1 \le l < i} X_l) \cap X_i$, one has that $j > i$ so that $A \in \partial_\theta X_j$. Since $X_j \in \mathcal{S}_{\theta'}$ and α_θ enjoys the running intersection property, there exists $h < k$ for which $\partial_\theta X_j = X_h \cap X_j$. To sum up, one has $A \in \partial_\theta X_j = X_h \cap X_j$ so that $A \in X_h$. Finally, since $h < k \le i$, one has $A \in X_h \cap X_i \subseteq \partial_\theta X_i$ so that $A \in \partial_\theta X_i$ (contradiction). □

The following lemma provides a formula for $[\theta']_{\mathbf{f}}$.

Lemma 10.3. Let θ be a canonical expression, and let $\alpha_\theta = (X_1, \ldots, X_n)$. Let θ' be a subexpression of θ, and let $\alpha_{\theta'} = (X_k, \ldots, X_m)$ for some k and m, $1 \le k \le m \le n$. Let $(\mathcal{R}_{\theta'}, \mathcal{S}_{\theta'})$ be the bipartition of $\mathcal{H}_{\theta'}$ defined by eq. (20), and let $I_{\theta'} = \{i : X_i \in \mathcal{R}_{\theta'}\}$ and $J_{\theta'} = \{i : X_i \in \mathcal{S}_{\theta'}\}$. If $\mathbf{f} = (f_1, \ldots, f_n)$ is a valid interpretation of θ, then

$$[\theta']_{\mathbf{f}} = \frac{\prod_{k \le i \le m} f_i}{p_{\theta'} \times \prod_{i \in I_{\theta'}} f_i^{\downarrow \partial_\theta X_i}} \tag{21}$$

where $p_{\theta'}$ is a function of $\mathrm{COM}(\mathcal{S}_{\theta'})$ if $\mathrm{COM}(\mathcal{S}_{\theta'}) \neq \emptyset$, and is a constant otherwise.

Proof. We now prove the statement by induction on the cardinality of $\mathcal{H}_{\theta'}$.
BASIS. If $k = m$, then $[\theta']_{\mathbf{f}} = f_k$. On the other hand, $\mathcal{H}_{\theta'} = \{X_k\}$ so that $\mathcal{R}_{\theta'} = \emptyset$, $\mathcal{S}_{\theta'} = \{X_k\}$ and $\mathrm{COM}(\mathcal{S}_{\theta'}) = \emptyset$. Therefore, eq. (21) holds with $p_{\theta'} = 1$.
INDUCTION. Assume that $m > k$ and let $\theta' = (\theta_1) \triangleright (\theta_2)$. Let $\alpha_{\theta_1} = (X_k, \ldots, X_l)$ and $\alpha_{\theta_2} = (X_{l+1}, \ldots, X_m)$. Consider the bipartitions of $\mathcal{H}_{\theta_1}$ and $\mathcal{H}_{\theta_2}$ defined by eq. (20):

$$\mathcal{R}_{\theta_1} = \{X_i \in \mathcal{H}_{\theta_1} : \exists\, j, k \le j < i,\ X_j \cap X_i = \partial_\theta X_i\} \qquad \mathcal{S}_{\theta_1} = \mathcal{H}_{\theta_1} - \mathcal{R}_{\theta_1}$$

$$\mathcal{R}_{\theta_2} = \{X_i \in \mathcal{H}_{\theta_2} : \exists\, j, l+1 \le j < i,\ X_j \cap X_i = \partial_\theta X_i\} \qquad \mathcal{S}_{\theta_2} = \mathcal{H}_{\theta_2} - \mathcal{R}_{\theta_2}.$$

Let

$$I_{\theta_h} = \{i : X_i \in \mathcal{R}_{\theta_h}\} \qquad J_{\theta_h} = \{i : X_i \in \mathcal{S}_{\theta_h}\} \qquad (h = 1, 2).$$

Thus, $I_{\theta_1} \cup J_{\theta_1} = \{k, \ldots, l\}$ and $I_{\theta_2} \cup J_{\theta_2} = \{l+1, \ldots, m\}$.

By the inductive hypothesis, one has

$$[\theta_h]_{\mathbf{f}} = \frac{\prod_{i \in I_{\theta_h} \cup J_{\theta_h}} f_i}{p_{\theta_h} \times \prod_{i \in I_{\theta_h}} f_i^{\downarrow \partial_\theta X_i}} \qquad (h = 1, 2)$$

where p_{θ_h} is a function of COM($\mathcal{S}_{\theta_h}$) if COM($\mathcal{S}_{\theta_h}$) $\neq \emptyset$, and is a constant otherwise, $h = 1, 2$. Then, one has

$$[\theta']_{\mathbf{f}} = [\theta_1]_{\mathbf{f}} \times \frac{[\theta_2]_{\mathbf{f}}}{\sum_{A \in V(\theta_2) - V(\theta_1)} [\theta_2]_{\mathbf{f}}}.$$

Explicitly, one has

$$[\theta']_{\mathbf{f}} = \frac{\prod_{k \leq i \leq m} f_i}{p_{\theta_1} \times p_{\theta_2} \times \sigma \times \prod_{i \in I_{\theta_1} \cup I_{\theta_2}} f_i^{\downarrow \partial_\theta X_i}} \tag{22}$$

where

$$\sigma = \sum_{A \in V(\theta_2) - V(\theta_1)} [\theta_2]_{\mathbf{f}} = \sum_{A \in V(\theta_2) - V(\theta_1)} \frac{\prod_{l+1 \leq i \leq m} f_i}{p_{\theta_2} \times \prod_{i \in I_{\theta_2}} f_i^{\downarrow \partial_\theta X_i}}. \tag{23}$$

Let J' be the subset of J_{θ_2} defined as follows:

$$J' = \{j \in J_{\theta_2} : \exists\, X_i \in \mathcal{H}_{\theta_1} \quad X_i \cap X_j = \partial_\theta X_j\}.$$

Then one has

$$I_{\theta'} = I_{\theta_1} \cup I_{\theta_2} \cup J' \qquad J_{\theta'} = J_{\theta_1} \cup J_{\theta_2} - J'.$$

We shall prove that there exists a function g such that

(i) $\sigma = g \times \prod_{j \in J'} f_j^{\downarrow \partial_\theta X_j}$;

(ii) $p_{\theta_1} \times p_{\theta_2} \times g$ is a function of COM($\mathcal{S}_{\theta'}$) if COM($\mathcal{S}_{\theta'}$) $\neq \emptyset$, and is a constant otherwise.

Then, by (i) and (ii), eq. (22) can be re-written as

$$[\theta']_{\mathbf{f}} = \frac{\prod_{k \leq i \leq m} f_i}{p_{\theta_1} \times p_{\theta_2} \times g \times \prod_{i \in I_{\theta_1} \cup I_{\theta_2} \cup J'} f_i^{\downarrow \partial_\theta X_i}}$$

which proves the statement with $p_{\theta'} = p_{\theta_1} \times p_{\theta_2} \times g$.

At this point, what remains to prove is the existence of a function g having properties (i) and (ii).

Proof of (i). First of all, observe that, for each set $X_i \in \mathcal{H}_{\theta_2}$ (that is, for each $i \in I_{\theta_2} \cup J_{\theta_2}$), since $X_i \cap V(\theta_1) \subseteq \partial_\theta X_i$ one has that $(X_i - \partial_\theta X_i) \cap V(\theta_1) = \emptyset$ so that $X_i - \partial_\theta X_i \subseteq V(\theta_2) - V(\theta_1)$. Let

$$U = \cup_{i \in I_{\theta_2}} (X_i - \partial_\theta X_i) \qquad W = \cup_{i \in J_{\theta_2}} (X_i - \partial_\theta X_i) \qquad V' = V(\theta_2) - (U \cup W).$$

Therefore, one has $U \cap V(\theta_1) = W \cap V(\theta_1) = \emptyset$ and

$$V(\theta_2) - V(\theta_1) = U \cup W \cup (V' - V(\theta_1)).$$

Consider now the summation $\sum_{A\in V(\theta_2)-V(\theta_1)}$ in eq. (23). We first sum out the variables in U, then the variables in W, and finally the variables in $V'-V(\theta_1)$. In other words, we break the summation $\sum_{A\in V(\theta_2)-V(\theta_1)}$ into

$$\sum_{A\in V'-V(\theta_1)} \quad \sum_{A\in W} \qquad \sum_{A\in U} \qquad .$$

Let us begin to sum out the variables in U. Let $I_{\theta_2}=\{i_1,\ldots,i_{s-1},i_s\}$ where $i_1<\ldots<i_{s-1}<i_s$; thus,

$$U=(X_{i_1}-\partial_\theta X_{i_1})\cup\ldots\cup(X_{i_{s-1}}-\partial_\theta X_{i_{s-1}})\cup(X_{i_s}-\partial_\theta X_{i_s}).$$

Then we sum out the variables in U in the following order: first the variables in $X_{i_s}-\partial_\theta X_{i_s}$, next the variables in $X_{i_{s-1}}-\partial_\theta X_{i_{s-1}}$, ..., last the variables in $X_{i_1}-\partial_\theta X_{i_1}$:

$$\sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_s}-\partial_\theta X_{i_s}} \frac{\left(\prod_{1\le r\le s} f_{i_r}\right)\times\left(\prod_{i\in J_{\theta_2}} f_i\right)}{p_{\theta_2}\times\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}}.$$

By part (a) of Lemma 10.2 for each r, $1\le r\le s$, one has $(X_{i_r}-\partial_\theta X_{i_r})\cap V(\mathcal{S}_{\theta_2})=\emptyset$ so that, since $\mathcal{S}_{\theta_2}=\{X_i : i\in J_{\theta_2}\}$ and p_{θ_2} is a function of $\mathrm{COM}(\mathcal{S}_{\theta_2})$ or a constant, we can move $\frac{\prod_{i\in J_{\theta_2}} f_i}{p_{\theta_2}}$ to the left of the leftmost summation $\sum_{A\in X_{i_1}-\partial_\theta X_{i_1}}$. Thus, we obtain:

$$\frac{\prod_{i\in J_{\theta_2}} f_i}{p_{\theta_2}} \times \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_s}-\partial_\theta X_{i_s}} \frac{\prod_{1\le r\le s} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}}.$$

Let

$$\sigma' = \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_s}-\partial_\theta X_{i_s}} \frac{\prod_{1\le r\le s} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}}.$$

We now prove that $\sigma'=1$. First of all, we re-write σ' as

$$\sigma' = \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_s}-\partial_\theta X_{i_s}} \frac{\prod_{1\le r\le s-1} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}} \times f_{i_s}.$$

Note that, since $X_{i_r}\cap X_{i_s}\subseteq \partial_\theta X_{i_s}$ for each $r<s$, one has that

$$X_{i_r}\cap(X_{i_s}-\partial_\theta X_{i_s})=\emptyset$$

so that we can move $\frac{\prod_{1\le r\le s-1} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}}$ to the left of the summation $\sum_{A\in X_{i_s}-\partial_\theta X_{i_s}}$. Thus, we have

$$\sigma' = \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_{s-1}}-\partial_\theta X_{i_{s-1}}} \left(\frac{\prod_{1\le r\le s-1} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}} \times \sum_{A\in X_{i_s}-\partial_\theta X_{i_s}} f_{i_s}\right)$$

$$= \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_{s-1}}-\partial_\theta X_{i_{s-1}}} \left(\frac{\prod_{1\le r\le s-1} f_{i_r}}{\prod_{1\le r\le s} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}} \times f_{i_s}^{\downarrow\partial_\theta X_{i_s}}\right)$$

$$= \sum_{A\in X_{i_1}-\partial_\theta X_{i_1}} \cdots \sum_{A\in X_{i_{s-1}}-\partial_\theta X_{i_{s-1}}} \frac{\prod_{1\le r\le s-1} f_{i_r}}{\prod_{1\le r\le s-1} f_{i_r}^{\downarrow\partial_\theta X_{i_r}}}.$$

By repeating the same argument for $r = s-1, \ldots, 1$, we obtain $\sigma' = 1$. So,

$$\sigma = \sum_{A\in V'-V(\theta_1)} \sum_{A\in W} \frac{\prod_{i\in J_{\theta_2}} f_i}{p_{\theta_2}}.$$

We now sum out the variables in W. By part (b) of Lemma 10.2, for each $i \in J_{\theta_2}$, each variable in $X_i - \partial_\theta X_i$ is unique in $\mathcal{S}_{\theta_2}$. Therefore, one has

$$(X_i - \partial_\theta X_i) \cap \mathrm{COM}(\mathcal{S}_{\theta_2}) = \emptyset$$

so that we can move $\frac{1}{p_{\theta_2}}$ to the left of $\sum_{A\in W}$; moreover, owing to the uniqueness of the variables in W, we have

$$\sum_{A\in W} \prod_{i\in J_{\theta_2}} f_i = \prod_{i\in J_{\theta_2}} f_i^{\downarrow\partial_\theta X_i}.$$

To sum up, we have

$$\sum_{A\in W} \frac{\prod_{i\in J_{\theta_2}} f_i}{p_{\theta_2}} = \frac{1}{p_{\theta_2}} \sum_{A\in W} \prod_{i\in J_{\theta_2}} f_i = \frac{\prod_{i\in J_{\theta_2}} f_i^{\downarrow\partial_\theta X_i}}{p_{\theta_2}}$$

and hence

$$\sigma = \sum_{A\in V'-V(\theta_1)} \frac{\prod_{i\in J_{\theta_2}} f_i^{\downarrow\partial_\theta X_i}}{p_{\theta_2}}. \tag{24}$$

Note that, since $(X_i - \partial_\theta X_i) \cap \mathrm{COM}(\mathcal{S}_{\theta_2}) = \emptyset$ for each $i \in J_{\theta_2}$, one has $\mathrm{COM}(\mathcal{S}_{\theta_2}) \subseteq V'$. Consider now the factors $f_j^{\downarrow\partial_\theta X_j}$ for $j \in J' \subseteq J_{\theta_2}$. By the very definition of J' one has that $\partial_\theta X_j \subseteq V(\theta_1)$ for each $j \in J'$, so that in eq. (24) we can move the factor $f_j^{\downarrow\partial_\theta X_j}$ to the left of the summation $\sum_{A\in V'-V(\theta_1)}$:

$$\sigma = \prod_{j\in J'} f_j^{\downarrow\partial_\theta X_j} \times \sum_{A\in V'-V(\theta_1)} \frac{\prod_{i\in J_{\theta_2}-J'} f_i^{\downarrow\partial_\theta X_i}}{p_{\theta_2}}$$

which with

$$g = \sum_{A\in V'-V(\theta_1)} \frac{\prod_{i\in J_{\theta_2}-J'} f_i^{\downarrow\partial_\theta X_i}}{p_{\theta_2}}$$

reduces to the form stated in (i).

Proof of (*ii*). Recall that p_{θ_h} is a function of $\mathrm{COM}(\mathcal{S}_{\theta_h})$ if $\mathrm{COM}(\mathcal{S}_{\theta_h}) \neq \emptyset$, and is a constant otherwise, $h = 1, 2$. Moreover, g is a function of the set $V' \cap V(\theta_1)$ which is equal to the union of $\mathrm{COM}(\mathcal{S}_{\theta_2}) \cap V(\theta_1)$ with $\mathrm{UNI}(\mathcal{S}_{\theta_2}) \cap V(\theta_1)$. Finally, since

$$\mathrm{COM}(\mathcal{S}_{\theta'}) = \mathrm{COM}(\mathcal{S}_{\theta_1}) \cup \mathrm{COM}(\mathcal{S}_{\theta_2}) \cup (\mathrm{UNI}(\mathcal{S}_{\theta_2}) \cap V(\theta_1)),$$

$p_{\theta_1} \times p_{\theta_2} \times g$ is a function of $\mathrm{COM}(\mathcal{S}_{\theta'})$ if $\mathrm{COM}(\mathcal{S}_{\theta'}) \neq \emptyset$, and is a constant otherwise. □

The following is an illustrative example of eq. (21).

Example 10.4. Consider again the canonical expression θ of Example 10.1. Let $\mathbf{f} = (f_1, f_2, \ldots, f_8)$ be a valid interpretation of θ, and let f_i be the distribution function on X_i, $1 \leq i \leq 8$, where

$$X_1 = ABCD \qquad X_2 = ABE \qquad X_3 = BCF \qquad X_4 = FL$$

$$X_5 = CDG \qquad X_6 = ADH \qquad X_7 = CI \qquad X_8 = BCM$$

We now show that eq. (21) holds for the three subexpressions $\theta_1 = (ABE \rhd BCF) \rhd FL$, $\theta_2 = (CDG \rhd ADH) \rhd (CI \rhd BCM)$ and $\theta_3 = (\theta_1) \rhd (\theta_2)$ of θ mentioned in Example 10.1.

- Recall that $\mathcal{S}_{\theta_1} = \{ABE, BCF\}$ and $\mathrm{COM}(\mathcal{S}_{\theta_1}) = B$; moreover, $\mathcal{R}_{\theta_1} = \{X_4 = FL\}$ so that $I_{\theta_1} = \{4\}$. For $[\theta_1]_{\mathbf{f}}$ one has

$$[\theta_1]_{\mathbf{f}} = \left(f_2 \times \frac{f_3}{f_3^{\downarrow B}}\right) \times \frac{f_4}{f_4^{\downarrow F}} = \frac{f_2 \times f_3 \times f_4}{f_3^{\downarrow B} \times f_4^{\downarrow F}}$$

  which reduces to eq. (21) with

  - $p_{\theta_1} = f_3^{\downarrow B}$, which is a function of $\mathrm{COM}(\mathcal{S}_{\theta_1})$;
  - $\prod_{i \in I_{\theta_1}} f_i^{\downarrow \partial_\theta X_i} = f_4^{\downarrow F}$ which is the marginal of f_4 on $\partial_\theta X_4 = F$.

- Recall that $\mathcal{S}_{\theta_2} = \{CDG, ADH, BCM\}$ and $\mathrm{COM}(\mathcal{S}_{\theta_2}) = CD$; moreover, $\mathcal{R}_{\theta_2} = \{X_7 = CI\}$ so that $I_{\theta_2} = \{7\}$. For $[\theta_2]_{\mathbf{f}}$ one has

$$[\theta_2]_{\mathbf{f}} = \frac{f_5 \times f_6}{f_6^{\downarrow D}} \times \frac{\frac{f_7 \times f_8}{f_8^{\downarrow C}}}{\sum_{B,I,M} \frac{f_7 \times f_8}{f_8^{\downarrow C}}} = \frac{f_5 \times f_6 \times f_7 \times f_8}{f_6^{\downarrow D} \times f_7^{\downarrow C} \times f_8^{\downarrow C}}$$

  which reduces to eq. (21) with

  - $p_{\theta_2} = f_6^{\downarrow D} \times f_8^{\downarrow C}$, which is a function of $\mathrm{COM}(\mathcal{S}_{\theta_2})$;
  - $\prod_{i \in I_{\theta_2}} f_i^{\downarrow \partial_\theta X_i} = f_7^{\downarrow C}$.

- Recall that $\mathcal{S}_{\theta_3} = \{ABE, BCF, CDG, ADH\}$ and $\mathrm{COM}(\mathcal{S}_{\theta_3}) = ABCD$; moreover, $\mathcal{R}_{\theta_3} = \{X_4 = FL, X_7 = CI, X_8 = BCM\}$ so that $I_{\theta_3} = \{4, 7, 8\}$. For $[\theta_3]_{\mathbf{f}}$ one has

$$[\theta_3]_{\mathbf{f}} = \frac{f_2 \times f_3 \times f_4}{f_3^{\downarrow B} \times f_4^{\downarrow F}} \times \frac{\frac{f_5 \times f_6 \times f_7 \times f_8}{f_6^{\downarrow D} \times f_7^{\downarrow C} \times f_8^{\downarrow C}}}{\sum_{D,G,I,H,M} \frac{f_5 \times f_6 \times f_7 \times f_8}{f_6^{\downarrow D} \times f_7^{\downarrow C} \times f_8^{\downarrow C}}}$$

$$= \frac{f_2 \times f_3 \times f_4 \times f_5 \times f_6 \times f_7 \times f_8}{f_3^{\downarrow B} \times f_4^{\downarrow F} \times f_6^{\downarrow D} \times f_7^{\downarrow C} \times f_8^{\downarrow BC} \times \sum_D \frac{f_5^{\downarrow CD} \times f_6^{\downarrow AD}}{f_6^{\downarrow D}}}$$

which reduces to eq. (21) with

  - $p_{\theta_3} = f_3^{\downarrow B} \times f_6^{\downarrow D} \times \sum_D \frac{f_5^{\downarrow CD} \times f_6^{\downarrow AD}}{f_6^{\downarrow D}}$, which is a function of $\mathrm{COM}(\mathcal{S}_{\theta_3})$;
  - $\prod_{i \in I_{\theta_3}} f_i^{\downarrow \partial_\theta X_i} = f_4^{\downarrow F} \times f_7^{\downarrow C} \times f_8^{\downarrow BC}$.

At this point, we are in a position to prove eq. (17).

Theorem 10.5. Let θ be a canonical expression and let $\alpha_\theta = (X_1, \ldots, X_n)$ be its base sequence. If $\mathbf{f} = (f_1, \ldots, f_n)$ is a valid interpretation of θ, then

$$[\theta]_{\mathbf{f}} = f_1 \times \prod_{2 \le i \le n} \frac{f_i}{f_i^{\downarrow \partial_\theta X_i}}.$$

Proof. Since θ is a subexpression of itself, we can apply Lemma 10.3 with $\theta' = \theta$. Since $\mathcal{R}_\theta = \{X_2, \ldots, X_n\}$ and $\mathcal{S}_\theta = \{X_1\}$, one has $I_\theta = \{2, \ldots, n\}$, $J_\theta = \{1\}$ and $\mathrm{COM}(\mathcal{S}_\theta) = \emptyset$, so that eq. (21) reduces to

$$[\theta]_{\mathbf{f}} = \frac{\prod_{1 \le i \le n} f_i}{p_\theta \times \prod_{2 \le i \le n} f_i^{\downarrow \partial_\theta X_i}}$$

where p_θ is a constant. What remains to prove is that $p_\theta = 1$. Since $[\theta]_{\mathbf{f}}$ is an extension of f_1, one has

$$\sum_{A \in V(\theta) - X_1} [\theta]_{\mathbf{f}} = f_1.$$

On the other hand, it is easy to see that

$$\sum_{A \in V(\theta) - X_1} [\theta]_{\mathbf{f}} = \sum_{A \in X_2 - \partial_\theta X_2} \cdots \sum_{A \in X_n - \partial_\theta X_n} \frac{\prod_{1 \le i \le n} f_i}{p_\theta \times \prod_{2 \le i \le n} f_i^{\downarrow \partial_\theta X_i}}$$

$$= \sum_{A \in X_2 - \partial_\theta X_2} \cdots \sum_{A \in X_{n-1} - \partial_\theta X_{n-1}} \frac{\prod_{1 \le i \le n-1} f_i}{p_\theta \times \prod_{2 \le i \le n-1} f_i^{\downarrow \partial_\theta X_i}} = \ldots = \frac{f_1}{p_\theta}$$

which implies that $p_\theta = 1$. □

ACKNOWLEDGEMENT

The author is grateful to an anonymous reviewer whose comments were helpful to improve the organization and readability of the paper.

(Received October 25, 2013)

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] J. R. S. Blair and B.W. Peyton: An Introduction to Chordal Graphs and Clique Trees. Technical Report ORNL/TM-12203, 1992.

[3] I. Csiszár: I-divergence geometry of probability distributions and minimization problems. Ann. Probab. *3* (1975), 146–158.

[4] A. P. Dawid: Applications of a general propagation algorithm for probabilistic expert systems. Statist. Comput. *2* (1992), 25–36.

[5] W. E. Deming and F. F. Stephan: On least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann. Math. Stat. *11* (1940), 427–444.

[6] H. Hara and A. Takemura: Boundary cliques, clique trees and perfect sequences of maximaal cliques of a chordal graph. arXiv:cs/0607055v1 [cs.DM], 11 July 2006, pp. 1–24.

[7] R. Jiroušek: Composition of probability measures on finite spaces. In: Proc. XIII International Conf. on Uncertainty in Artificial Intelligence (D. Geiger and P. P. Shenoy, eds.), Morgan Kaufmann, San Francisco 1997, pp. 274–281.

[8] R. Jiroušek: Decomposition of multidimensional distributions represented by perfect sequences. Ann. Math. Artif. Intelligence *5* (2002), 215–226.

[9] R. Jiroušek: Detection of independence relations from persegrams. In: Proc. VI International Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Annecy 2002, Vol. C, pp. 1261–1267.

[10] R. Jiroušek: Persegrams of compositional models revisited: conditional independence. In: Proc. XII International Conf. on Information Processing and Management of Uncertainty in Knowledge-Based Systems (L. Magdalena, M. Ojeda-Aciego and J. L. Verdegay, eds.), Malaga 2008, pp. 915–922.

[11] R. Jiroušek: Foundations of compositional model theory. Internat. J. General Systems *40* (2011), 623–678.

[12] R. Jiroušek: Local computations in Dempster–Shafer theory of evidence. Internat. J. Approx. Reason. *53* (2012), 1155–1167.

[13] R. Jiroušek and P. P. Shenoy: Compositional models in valuation-based systems. Internat. J. Approx. Reason. *55* (2014), 277–293.

[14] R. Jiroušek and J. Vejnarová: General framework for multidimensional models. Internat. J. Approx. Reas. *18* (2003), 107–127.

[15] R. Jiroušek, J. Vejnarová. and M. Daniels: Composition models of belief functions. In: Proc. V Symp. on Imprecise Probabilities and Their Applications (G. De Cooman, J. Vejnarová, and M. Zaffalon, eds.), Action M Agency, Prague 2007, pp. 243–252.

[16] V. Kratochvíl: Characteristic properties of equivalent structures in compositional models. Internat. J. Approx. Reason. *52* (2011), 599–612.

[17] V. Kratochvíl: Probabilistic compositional models: solution of an equivalence problem. Internat. J. Approx. Reason. *54* (2013), 590–601.

[18] S. L. Lauritzen: Graphical Models. Oxford University Press, Oxford 1996.

[19] H.-J. Lenz and B. Talheim: A formal framework of aggregation for the OLAP-OLTP model. J. of Universal Computer Science *15* (2009), 273–303.

[20] F. M. Malvestuto: Tree and local computations in a cross-entropy minimization problem with marginal constraints. Kybernetika *46* (2010), 621–654.

[21] F. M. Malvestuto: The sum-product algorithm: algebraic independence and computational aspects. Kybernetika *49* (2013), 4–22.

[22] F. M. Malvestuto: A join-like operator to combine data cubes, and answer queries from multiple data cubes. Unpublished manuscript, 2013.

[23] F. M. Malvestuto and E. Pourabbas: Local computation of answers to table queries on summary databases. In: Proc. XVII International Conference on Scientific and Statistical Database Management, Santa Barbara 2005, pp. 263–270.

[24] F. Mosteller: On pooling data. J. Amer. Statist. Assoc. *43* (1948), 231–242.

[25] J. Pearl: Probabilistic Reasoning in Intelligent Systems. Morgan Kaufmann Pub., San Mateo 1988.

[26] E. Pourabbas and A. Shoshani: Efficient estimation of joint queries from multiple OLAP databases. ACM Trans. Database Systems *32* (2007), 1, Article No. 2.

[27] E. Pourabbas and A. Shoshani: Improving estimation accuracy of aggregate queries on data cubes. Data and Knowledge Engrg. *69* (2010), 50–72.

[28] P. P. Shenoy and G. Shafer: Axioms for probability and belief-function propagation. In: Uncertainty in Artificial Intelligence (R. D. Shachter, T. Levitt, J. F. Lemmer, and L. N. Kanel, eds.), North-Holland 1990, Vol. 4.

[29] R. E. Tarjan and M. Yannakakis: Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. SIAM J. Comput. *13* (1984), 566–579.

[30] V. Verma, F. Gagliardi, and C. Ferretti: On Pooling Data and Measures. Working Paper No. 84 University of Siena, 2009.

[31] J. Vomlel: Integrating inconsistent data in a probabilistic model. J. Appl. Non-Classical Logics *49* (2003), 4–22.

Francesco M. Malvestuto, Department of Computer Science, Sapienza University of Rome, Via Salaria 113, 00198 Rome. Italy.

e-mail: malvestuto@di.uniroma1.it