Theoretical Computer Science
81, 117–126 (1991)

Abstract functional dependency structures

František Matúš

Institute of Information Theory and Automation
Czechoslovak Academy of Sciences
Pod vodárenskou věží 4, 182 08 Prague
Czechoslovakia

Abstract

The functional dependency (FD) structures known mainly from the database relationships theory are appeared to comprehend other kinds of dependencies among objects like functions, boolean or random variables, etc. We examine the representations of FD-relations in the matroid theory manner. A Galois connection of the FD-relations with closure systems is elaborated.

INTRODUCTION

The notion of "functional dependence" seems to be of the most fundamental ones in mathematics and applications. We are going to touch it in the way indicated by the Definition below which covers all presented situations and enables to analyse their common features.

Let the letter N be reserved in all paper for a fixed finite set. The power set of N is denoted by $\mathcal{P}(N)$, the set of all ordered pairs (I, J) where I, J are subsets of N by $\mathcal{Q}(N)$ and the set of all equivalences on N by $\mathcal{E}(N)$.

Definition

A set $\mathcal{N} \subset \mathcal{Q}(N)$ is called FD-relation on N if and only if it fulfils these three properties (we read $(I, J) \in \mathcal{N}$ as J depends functionally on I)

1. $N \supset I \supset J \;\Rightarrow\; (I, J) \in \mathcal{N}$

2. $(I, J) \in \mathcal{N},\ (J, K) \in \mathcal{N} \;\Rightarrow\; (I, K) \in \mathcal{N}$

3. $(I, J) \in \mathcal{N},\ (I, K) \in \mathcal{N} \;\Rightarrow\; (I, J \cup K) \in \mathcal{N}$

Before outlining the content of the paper we want to stress simple connections of FD-relations, closure operators and closure systems.

Let us define the mapping $c_{\mathcal{M}} : \mathcal{P}(N) \to \mathcal{P}(N)$ for $\mathcal{M} \subset \mathcal{Q}(N)$ by the equality

$$c_{\mathcal{M}}(I) = \bigcup \{J \subset N; (I, J) \in \mathcal{M}\}, \qquad I \subset N.$$

If $\mathcal{M} \in \mathbf{N}$ (the last symbol denotes the set of all FD-relations on N) it is easy to check that $c_{\mathcal{M}}$ is a closure operator on N (see [4] or [9])

$$I \subset J \subset N \ \Rightarrow\ I \cup c(I) \cup c(c(I)) \subset c(J)\ .$$

(we dropped index $\mathcal{M}$ here). Contrariwise, having a closure operator c on N the relation $\mathcal{M}_c \subset \mathcal{Q}(N)$ given by

$$\mathcal{M}_c = \{(I, J); J \subset c(I)\}$$

is evidently an FD-relation on N. From $\mathcal{M} = \mathcal{M}_{c_{\mathcal{M}}}$, $\mathcal{M}$ - an FD-relation on N, and $c = c_{\mathcal{M}_c}$, c - a closure operator on N, one can conclude that there is a one-to-one and onto correspondence between the set $\mathbf{N}$ of all FD-relations on N and the family of all closure operators on N.

It is valuable to remind in this connection the well known one-to-one onto correspondence between the family of all closure operators on N and the set $\mathbf{C}$ of all closure systems on N. Having a closure operator c the family $\mathcal{C} = \{I \subset N; c(I) = I\}$ (of all closed sets of c) is the closure system on N, i.e. it holds

$$[\forall \alpha \in A \quad I_\alpha \in \mathcal{C}] \ \Rightarrow\ \bigcap \{I_\alpha; \alpha \in A\} \in \mathcal{C}$$

(if $A = \emptyset$ we get $N \in \mathcal{C}$) and, on the other hand, to a closure system $\mathcal{C} \subset \mathcal{P}(N)$ the mapping: $c(I) = \bigcap \{J; I \subset J \in \mathcal{C}\}$ is a closure operator on N (cf. [4], [5]).

It seems that FD-relations arising from databases occured for the first time in the database theory twenty years ago (see e.g. [1], [6], [8]). Armstrong [1] first recognized that, in his terminology, the full families of dependencies would be treated separately from database relationships (the properties 1., 2. and 3. or a set of another ones equivalent with them, are refered to as Armstrong's axioms) and straightforward generalization leads to the FD-relations. Ullman [19] presents an elementary treatment of the functional dependence in database context, a history of the subject and a large bulk of references. From the latest works it would be mentioned those of Beeri et al. [3] and Saxena and Tripathi [16], [17]. Fagin reveals [8] that the functional dependences in databases have close connection to the logical dependences of boolean variables. Hence, FD-relations may arise in a Boolean algebras context. Following the statistical interpretation of databases due to Malvestuto [10] one may expect that also dependences among random variables are covered by this notion. We shall have possibility to discuss all these situations.

We can now describe more exactly the two parts into which our paper is divided. As may be noticed we keep Armstrong's direction and prefer an algebraic treatment of the subject like in matroid theory to the logical one used usually in databases; thus we analyse FD-relations on abstract level irrelevant to situations where they arise. We remark that the matroids, in other words the linear dependence relations, can be viewed as special FD-relations, namely those the closure operator of which has the exchange property (see [20])

4. $i, j \in N,\ I \subset N,\ i \notin c(I),\ i \in c(I \cup \{j\}) \ \Rightarrow\ j \in c(I \cup \{i\}).$

We start with a presentation of five examples giving rise to FD-relations. First one presents a unified approach to the following three and gives a common look at some problems of [1], [8], [10] and [14]. Representations of FD-relations in every from five situations are analysed similarly like the vector representations of matroids. This way of examination enables at once to solve the problems called in database theory "soundness" and "strong completeness".

In the second part of the paper a theoretical result (Theorem) on FD-relations is proved. Due to the one-to-one and onto correspondence between $\mathbf{N}$ and $\mathbf{C}$ enlightened above it can be seen that a theory of abstract functional dependence structures is nothing but the theory of closure operators and closure systems - classical algebraical and topological objects. Theorem states, loosely speaking, that the correspondence $\mathbf{N} \leftrightarrow \mathbf{C}$ can be derived also by a Galois connection induced by a natural binary relation (which is a subset of the Cartesian product $\mathcal{Q}(N) \times \mathcal{P}(N)$). Among the consequencies of it there are some fundamental results of Delobel, Casey [6] and Armstrong [1] (Th. 3., 4. and 7.).

EXAMPLES

We have collected here some situations in which one can naturally meet with the FD-relations.

Example 1

Let $(Z, \wedge)$ be a lower semi-lattice with the greatest element 1 and let $z = (z_i)_{i \in N}$ be a system of its elements indexed by N. We set $(I, J \subset N)$

$$(I, J) \in \mathcal{N}_z^{(1)} \;\Leftrightarrow\; \bigwedge_{i \in I} z_i \leq \bigwedge_{j \in J} z_j = z_J$$

(for $I = \emptyset$ one takes $z_I = 1$). Belonging of a pair (I, J) to $\mathcal{N}_z^{(1)}$ could be freely interpreted as the "logical dependence" of z_J on z_I (this terminology has its origin in the fact that if $(Z, \wedge)$ is the semi-lattice of a Boolean algebra $(Z, \wedge, \vee, ')$ then $z_I \leq z_J$ is sometimes written in the form $z_I \Rightarrow z_J$, see [4].

The following three examples are special cases of the foregoing one.

Example 2

Let to every $i \in N$ a nonempty set X_i be given and $x_i \in X_i$. We denote by $X_I = \prod_{i \in I} X_i, I \subset N$, the Cartesian product of the sets $X_i, i \in I$ ($X_\emptyset$ is supposed to be a fixed singleton, e.g. $X_\emptyset = \{\emptyset\}$), and if $x = (x_i)_{i \in N} \in X_N = X$ by x_I the canonical projection of $x \in X$ on X_I, i.e. $x_I = (x_i)_{i \in I}$ $(x_\emptyset = \emptyset)$. Having given an N-ary relation $A \subset X$ we write

$$(I, J) \in \mathcal{N}_A^{(2)} \;\Leftrightarrow\; [\forall x, y \in A : x_I = y_I \Rightarrow x_J = y_J]$$

(an equivalent requirement is that the binary relation $\{(x_I, x_J); x \in A\}$ is a function). This is equivalent to the usual formulation of the functional dependence statements from relational databases.

Using the notation $\tau_I^A = \{(x, y) \in A^2; x_I = y_I\}$ we see that $(I, J) \in \mathcal{N}_A^{(2)}$ means $\tau_I^A \subset \tau_J^A$. Thus, we become the same situation like in Example 1; namely the role of the semi-lattice $(Z, \wedge)$ plays here the set $\mathcal{E}(A)$ of all equivalences on A with the operation of intersection. The system z of elements of $\mathcal{E}(A)$ is given here by $z = \left(\tau_i^A\right)_{i\in N}$ (the superfluous braces at i were of course omitted); note that $\tau_I = \bigcap_{i\in I} \tau_i$, $\tau_\emptyset = A^2$.

Example 3

Let (for $i \in N$) f_i be a mapping defined on a nonempty set B with values in a set C and $f_I, I \subset N$, be the mapping of B into $C^I (C^\emptyset = \{\emptyset\})$ having in $b \in B$ the value $(f_i(b))_{i\in I} \in C^I$. Dependence of the functions in the family $f = (f_i)_{i\in N}$ can be described as (cf. [14])

$$(I, J) \in \mathcal{N}_f^{(3)} \;\Leftrightarrow\; \exists g_J^I : C^I \to C^J \text{ s.t. } f_J = g_J^I f_I$$

Assuming that ρ_i is the equivalence on B equal to the composition of the relations f_i and f_i^{-1}, i.e. $(b', b'') \in \rho_i \;\Leftrightarrow\; f_i(b') = f_i(b'')$, one can see

$$(I, J) \in \mathcal{N}_f^{(3)} \;\Leftrightarrow\; \bigcap_{i\in I} \rho_i = \rho_I \subset \rho_J.$$

Hence, this example is also a special case of Example 1 with the semi-lattice $(Z, \wedge)$ being the semi-lattice $(\mathcal{E}(B), \cap)$ like in Example 2 and $z = (\rho_i)_{i\in N}$.

Example 4

Let $\xi_i, i \in N$, be, for simplicity, a real random variable on the probability space $(\Omega, \mathbf{S}, P)$ and $\xi_I, I \subset N$, be the random vector $(\xi_i)_{i\in I}$ ($\xi_\emptyset$ is supposed to be equal identically to $\emptyset$), $\xi_N = \xi$. In chime with the foregoing example the "strong, or functional dependence" of the random variables creating ξ can be captured by the relation $\mathcal{N}_\xi^{(4)} \subset \mathcal{Q}(N)$ given by (cf. [10])

$$(I, J) \in \mathcal{N}_\xi^{(4)} \;\Leftrightarrow\; \exists\; g_J^I : \left(R^I, \mathcal{B}^I\right) \to \left(R^J, \mathcal{B}^J\right) \text{ s.t. } \xi_j = g_J^I \xi_I \qquad P - \text{a.s.}$$

where g_J^I is a Borel measurable function of R^I in $R^J, R^\emptyset = \{\emptyset\}$, and $\mathcal{B}^I$ is the family of all Borel sets in R^I (for the terminology see [18]).

More generally, for sub-σ-algebras of $\mathbf{S}$ the situation is as follows. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two sub-σ-algebras of $\mathbf{S}$ and $\mathcal{T}_1 \wedge \mathcal{T}_2$ be the smallest sub-σ-algebra of $\mathbf{S}$ containing $\mathcal{T}_1$ and $\mathcal{T}_2$. Consider the family Z of all complete sub-σ-algebras of $\mathbf{S}$ ($\mathcal{T}$ is complete with respect to P if and only if $E \in \mathcal{T},\; F \in \mathbf{S},\; P(E \triangle F) = 0 \;\Rightarrow\; F \in \mathcal{T}$; note that this is not the completeness in Lebesgue sense) and observe that $(Z, \wedge)$ is a semi-lattice with the greatest element

$$\mathcal{S}_\emptyset = \{E \in \mathbf{S}; P(E) \in \{0, 1\}\}.$$

Let $\mathcal{S}_i \in Z$, for $i \in N, \mathcal{S}_I = \bigwedge_{i\in I} \mathcal{S}_i$, for $I \subset N$, and $\mathcal{S} = (\mathcal{S}_i)_{i\in N}$. We set

$$(I, J) \in \mathcal{N}_\mathcal{S}^{(4)} \;\Leftrightarrow\; \mathcal{S}_I \le \mathcal{S}_J \Leftrightarrow \mathcal{S}_I \supset \mathcal{S}_J.$$

Evidently, if $\mathcal{S}_i = \xi_i^{-1}(\mathcal{B}) \wedge \mathcal{S}_\emptyset$ is the completion of the inverse image $\xi_i^{-1}(\mathcal{B})$ of the family $\mathcal{B}$ of all Borel sets in the real line under the mapping ξ_i, then the definition above accords with the previous one $\left(\mathcal{N}_\mathcal{S}^{(4)} = \mathcal{N}_\xi^{(4)}\right)$. In the case of finite probability spaces one can consider instead of $\mathcal{S}_i$ the corresponding equivalences on Ω and get the same situation like in Examples 2 and 3.

Example 5

As the matroid theory is closely connected to the theory of submodular set functions (the rank function of a matroid is submodular) one can expect that the FD-relations do too. We recall that a real function $r : \mathcal{P}(N) \to R$ is nondecreasing if $I \subset J \subset N \;\Rightarrow\; r(I) \leq r(J)$ and submodular if $I, J \subset N \;\Rightarrow\; r(I) + r(J) \geq r(I \cup J) + r(I \cap J)$ takes place, resp. It is easy to verify that the expression

$$(I, J) \in \mathcal{N}_r^{(5)} \;\Leftrightarrow\; r(I) = r(I \cup J)$$

defines in the case of any nondecreasing and submodular function r an FD-relation on N.

The last sentence serves, of course, as an indication that all five relations $\mathcal{N}_{\cdot}^{(\cdot)}$ presented in this paragraph are FD-relations. We do not supply proofs because of their triviality (the assertions of this kind are in database context commented like the "soundness" of Armstrong's axioms).

To illustrate the notion of the FD-relation we present Table reporting on the number of all FD-relations considered on an n-element set, $0 \leq n \leq 4$, (see second column) and on the number of types of the FD-relations (see third column); we say two FD-relations $\mathcal{N}, \mathcal{M}$ on N are of the same type if and only if one can be obtained from the other by a permutation π of the set N ($(I, J) \in \mathcal{N} \;\Leftrightarrow\; (\pi(I), \pi(J)) \in \mathcal{M}$.). The case $n = 4$ was verified only by performing a computer program.

REPRESENTATIONS

In every theory of dependency structures (see [2], [7], [11], [15]) there is highly important to investigate the adequacy of chosen models with respect to the situations leading to them. In matroid theory the representability of a given matroid by a family of vectors (in a linear space over a field) is up to now an attractive and not fully settled problem. In this spirit we shall prove here that any FD-relation can be represented by the attributes of every from described five examples (the "strong completeness" of Armstrong's axioms with respect to every example).

Proposition (Semi-lattice representations)

If $\mathcal{N}$ is any FD-relation on N then there exists a system $z = (z_i)_{i \in N}$ of elements of a (finite) semi-lattice $(Z, \wedge)$ such that $\mathcal{N} = \mathcal{N}_z^{(1)}$ (see Example 1).

Proof:

Let Z be the system of all closed sets of the closure operator $c_{\mathcal{N}} = c$ corresponding to the prescribed FD-relation $\mathcal{N}$ and for $I, J \in Z$ let $I \wedge J = c(I \cup J)$. Then $(Z, \wedge)$ is a finite semi-lattice (note that $I \wedge (J \wedge K) = (I \wedge J) \wedge K = c(I \cup J \cup K), I, J, K \in Z$) in which $I \leq J$ is equivalent to $I \supset J$. The greatest element in the semi-lattice is $c(\emptyset)$. We put $z_i = c(i)$ (we omit braces around i), $i \in N$, $z = (z_i)_{i \in N}$ and observe $c(I) = c\left(\bigcup_{i \in I} c(i)\right), I \subset N$. The following chain of equivalences encloses the proof:

$$(I, J) \in \mathcal{N}_z^{(1)} \;\Leftrightarrow\; \bigwedge_{i \in I} z_i \leq \bigwedge_{j \in J} z_j \;\Leftrightarrow\; c\left(\bigcup_{i \in I} c(i)\right) \supset c\left(\bigcup_{j \in J} c(j)\right) \;\Leftrightarrow$$
$$\Leftrightarrow\; c(I) \supset c(J) \;\Leftrightarrow\; c(I) \supset J \;\Leftrightarrow\; (I, J) \in \mathcal{N}.$$

$\square$

Remark 1 (Boolean representations)

First simple consequence of Proposition we want to mention is that the semi-lattice introduced there can be supposed, moreover, to be the semi-lattice of a Boolean algebra. In fact, to given $\mathcal{N} = \mathcal{N}_z^{(1)}$ is suffices only to consider the Boolean algebra $(\mathcal{P}(Z), \wedge, \vee, ')$ of all subsets of Z and the system $\tilde{z} = (\varphi(z_i))_{i \in N}$ of its elements, where φ is the injection of Z into $\mathcal{P}(Z)$ defined by $\varphi(x) = \{y \in Z; y \leq x\}$, $x \in Z$. Then the FD-relation $\mathcal{N}_{\tilde{z}}^{(1)}$ originated from the system $\tilde{z}$ in the semi-lattice $(\mathcal{P}(Z), \cap)$ of the Boolean algebra is evidently the same as $\mathcal{N}_z^{(1)}$ (due to the equality $\varphi(x \wedge y) = \varphi(x) \cap \varphi(y)$, $x, y \in Z$, the injection φ is isotone and we have

$$(I, J) \in \mathcal{N}_z^{(1)} \;\Leftrightarrow\; z_I \leq z_J \;\Leftrightarrow\; \varphi(z_I) = \bigcap_{i \in I} \varphi(z_i) = \tilde{z}_I \subset \tilde{z}_J \;\Leftrightarrow\; (I, J) \in \mathcal{N}_{\tilde{z}}^{(1)} \;).$$

Remark 2 (Relational representations)

The assertion that to every FD-relation $\mathcal{N}$ on N there are finite nonempty sets $X_i, i \in N$, and an N-ary relation $A \subset \prod_{i \in N} X_i$ holding $\mathcal{N} = \mathcal{N}_A^{(2)}$ (see Example 2) is known in the database theory as Armstrong completeness theorem (cf. [1], Th. 5). We indicate here that it is also a trivial consequence of Propositon.

Indeed, having $\mathcal{N} = \mathcal{N}_z^{(1)}$ we consider the semi-lattice $(\mathcal{E}(Z), \cap)$ and the system $\tau = (\tau_i)_{i \in N}$ of its elements, i.e. equivalences on Z, given by $\tau_i = \psi(z_i), i \in N$, where ψ is this injection of Z into $\mathcal{E}(Z)$

$$\psi(x) = \Big\{(x_1, x_2) \in Z^2; x_1 = x_2 \text{ or } (x_1 \leq x \text{ and } x_2 \leq x)\Big\}, \qquad x \in Z.$$

Knowing that $\psi(x \wedge y) = \psi(x) \cap \psi(y)$, $x, y \in Z$ we get

$$(I, J) \in \mathcal{N}_z^{(1)} \;\Leftrightarrow\; z_I \leq z_J \;\Leftrightarrow\; \tau_I \subset \tau_J.$$

Let now $X_i = Z/\tau_i$ be the partition of Z corresponding to the equivalence τ_i and for $x \in Z$ $[x]_i \subset Z$ denote the block from X_i containing x. We put $A = \chi(z) \subset \prod_{i \in N} X_i$, i.e. A is the image of Z by the mapping χ:

$$\chi(x) = ([x]_i)_{i \in N}, \quad x \in Z.$$

Finally

$$\begin{aligned}(I, J) \in \mathcal{N}_A^{(2)} \;&\Leftrightarrow\; \forall\, x, y \in Z\, [(\forall i \in I\ [x]_i = [y]_i) \Rightarrow (\forall j \in J\ [x]_j = [y]_j)] \\ &\Leftrightarrow\; \tau_I \subset \tau_J \Leftrightarrow (I, J) \in \mathcal{N}.\end{aligned}$$

We can conclude, that Proposition together with the standard embedding techniques (in Remark 1 we used the embedding φ of a semi-lattice into the lattice of all its subsets and in Remark 2 the embedding ψ into the lattice of equivalences on it) enable to find a common look at the notions of the logical and functional dependencies (cf. also Equivalence theorem of Fagin [8] which follows trivially from Remarks 1 and 2).

Remark 3 (Functional representations)

Every FD-relation can arise from a situation of Example 3. To see it we suppose $\mathcal{N} = \mathcal{N}_A^{(2)}$ for some $A \subset \prod_{i\in N} X_i$ according the foregoing remark and put $B = A, C = \bigcup_{i\in N} X_i$ and for $x = (x_i)_{i\in N} \in A$ $f_i(x) = x_i, i \in N$ (f_i is thus the i-th coordinate projection restricted on A). Immediately for $f = (f_i)_{i\in N}$ $\mathcal{N}_f^{(3)} = \mathcal{N}$ follows.

Remark 4 (Stochastic representations)

If $\mathcal{N}$ is arbitrary FD-relation then there is a system of (discrete) random variables $\xi = (\xi_I)_{i\in N}$ the functional dependence of which is described adequately by $\mathcal{N}$ (when one chooses $\Omega = A, \mathbf{S} = \mathcal{P}(A), P(\{x\}) > 0$ (arbitrary), $x \in A$, and $f = \xi$ Remark 4 is just a reformulation of Remark 3).

Remark 5

To any FD-relation $\mathcal{N}$ on N there exist nonnegative, nondecreasing and submodular function r on $\mathcal{P}(N)$ with the property $\mathcal{N} = \mathcal{N}_r^{(5)}$. In fact, under the notation of Remark 2 we set

$$r(I) = -\sum \left\{ \frac{|G|}{|A|} \ln \frac{|G|}{|A|} \; ; \; G \in A/\tau_I \right\}, \qquad I \subset N,$$

where $|G|$ denotes the cardinality of the block G from the partition of A corresponding to the equivalence τ_I. Hence, $r(I)$ is the entropy of ξ_I (cf.Remark 4). Instead of a verification that r has the desired properties we supply the precise reference (see [12] §§ 2.2 and 2.5).

We remark further that even an integer-valued function r may be found. Denoting by $q_\emptyset$ the function on $\mathcal{P}(N)$ given by $q_\emptyset(\emptyset) = 0$ and $q_\emptyset(I) = 1$, for $I \neq \emptyset$, one can easily see that every function $q_J, J \subset N$, given by $q_J(I) = q_\emptyset(I - J), I \subset N$, is nondecreasing and submodular. The choice $r = \sum_{K\in\mathcal{C}} q_K$ (the summation is here extended over all closed sets of the prescribed FD-relation $\mathcal{N}$) gives rise to $\mathcal{N}_r^{(5)} = \mathcal{N}$. This will be a clear consequence of Theorem from the next paragraph. Another view on these problems is in [20].

At the end of the section we formulate an open problem. Let an FD-relation $\mathcal{N}$ be called *binary* if and only if $\mathcal{N} = \mathcal{N}_A^{(2)}$ for some $A \subset \prod_{i\in N} X_i$ where all X_i are supposed to have exactly two elements. It would be interesting to know an axiomatic description of the class of all binary FD-relations.

GALOIS CONNECTION

The set $\mathbf{N} \subset \mathcal{P}(\mathcal{Q}(N))$ of all FD-relations on N and the set $\mathbf{C} \subset \mathcal{P}(\mathcal{P}(N))$ of all closure systems on N are also themselves closure systems (on $\mathcal{Q}(N)$ and $\mathcal{P}(N)$, resp.). So, the mapping $\mu : \mathcal{Q}(N) \to \mathcal{Q}(N)$ given by

$$\mu(\mathcal{M}) = \bigcap \{\mathcal{N} \in \mathbf{N}; \mathcal{N} \supset \mathcal{M}\}, \qquad \mathcal{M} \subset \mathcal{Q}(N),$$

is the closure operator on $\mathcal{Q}(N)$ and $\mathcal{N} \in \mathbf{N} \Leftrightarrow \mu(\mathcal{N}) = \mathcal{N}$. Similarly, the mapping $\delta : \mathcal{P}(N) \to \mathcal{P}(N)$,

$$\delta(\mathcal{D}) = \bigcap \{\mathcal{C} \in \mathbf{C}; \mathcal{C} \supset \mathcal{D}\}, \qquad \mathcal{D} \subset \mathcal{P}(N),$$

is the corresponding closure operator on $\mathcal{P}(N)$ and $\mathcal{C} \in \mathbf{C} \Leftrightarrow \delta(\mathcal{C}) = \mathcal{C}$. The sets $\mathbf{N}$ and $\mathbf{C}$ will be considered to be lattices (see [9], [5]). Let us mention that the closure operators μ and δ can be written in other "iterative" forms. If $\mathcal{M} \in \mathcal{Q}(N)$ and we set

$$\begin{aligned} \mathcal{M}_0 &= \mathcal{M} \cup \{(I,J); N \supset I \supset J\} \\ \mathcal{M}_{n+1} &= \mathcal{M}_n \cup \{(I, J \cup K); (I,J) \in \mathcal{M}_n, (I,K) \in \mathcal{M}_n\} \cup \\ & \cup \{(I,K); \exists J \subset N : (I,J) \in \mathcal{M}_n, (J,K) \in \mathcal{M}_n\}, \quad n \geq 0, \end{aligned}$$

then $\mu(\mathcal{M}) = \bigcup_{n \geq 0} \mathcal{M}_n$ (in the database terminology it is the set of all statements provable from $\mathcal{M}$ by Armstrong's axioms) and similarly

$$\begin{aligned} \mathcal{D}_0 &= \mathcal{D} \cup \{N\} \\ \mathcal{D}_{n+1} &= \mathcal{D}_n \cup \{K \cap L; K \in \mathcal{D}_n, L \in \mathcal{D}_n\}, \quad n \geq 0 \end{aligned}$$

and $\delta(\mathcal{D}) = \bigcup_{n \geq 0} \mathcal{D}_n$ (see [5], Ch II., § 5), where $\mathcal{D} \subset \mathcal{P}(N)$.

The main idea of this section is to introduce the binary relation $\mathbf{p} \subset \mathcal{Q}(N) \times \mathcal{P}(N)$ defined by $((I,J),K) \in \mathbf{p} \;\Leftrightarrow\; (I \subset K \Rightarrow J \subset K)$ (in [1] $\mathbf{p}$ occurs in a latent form). Then these two mappings

$$\begin{aligned} \mathcal{M} &\rightarrow \mathcal{M}^* = \{K \in \mathcal{P}(N)\ ;\ \forall (I,J) \in \mathcal{M} : ((I,J),K) \in \mathbf{p}\}, \quad \mathcal{M} \subset \mathcal{Q}(N), \\ \mathcal{D} &\rightarrow \mathcal{D}^* = \{(I,J) \in \mathcal{Q}(N)\ ;\ \forall K \in \mathcal{D} : ((I,J),K) \in \mathbf{p}\}, \quad \mathcal{D} \subset \mathcal{P}(N). \end{aligned}$$

form the Galois connection arising from $\mathbf{p}$ (cf. [4], Ch. 5, or [5] Ch. 2).

Theorem

The mappings $\mathcal{M} \rightarrow \mathcal{M}^{**}$ and $\mathcal{D} \rightarrow \mathcal{D}^{**}$ coincide with the closure operators μ and δ, respectively. The Galois connection $\mathcal{M} \rightarrow \mathcal{M}^*, \mathcal{D} \rightarrow \mathcal{D}^*$ establishes a pair of anti-isomorphisms between the lattices $\mathbf{N}$ and $\mathbf{C}$ each one being the inverse of the other.

Proof:

Let us start with these observations

$$\begin{aligned} ((I,J),N) &\in \mathbf{p}, \qquad I \subset N, J \subset N, \\ ((I,J),K) &\in \mathbf{p}, ((I,J),L) \in \mathbf{p} \;\Rightarrow\; ((I,J), K \cap L) \in \mathbf{p}, \end{aligned}$$

getting $\mathcal{M}^* \in \mathbf{C}$ for any $\mathcal{M} \subset \mathcal{Q}(N)$ and quite analogically, the observations

$$\begin{aligned} ((I,J),K) &\in \mathbf{p}, \qquad J \subset I \subset N, K \subset N, \\ ((I,J),L) &\in \mathbf{p}, ((J,K),L) \in \mathbf{p} \Rightarrow ((I,K),L) \in \mathbf{p}, \\ ((I,J),L) &\in \mathbf{p}, ((I,K),L) \in \mathbf{p} \Rightarrow ((I, J \cup K), L) \in \mathbf{p}, \end{aligned}$$

imply $\mathcal{D}^* \in \mathbf{N}, \mathcal{D} \subset \mathcal{P}(N)$.

We are going now to prove that $\mathcal{N}^{**} = \mathcal{N}$ holds for any FD-relation on N. In one direction ($\mathcal{N}^{**} \supset \mathcal{N}$) it is trivial, namely $\mathcal{M}^{**} \supset \mu(\mathcal{M})$ for $\mathcal{M} \subset \mathcal{Q}(N)$ takes place. On the other hand let $(I,J) \in \mathcal{N}^{**}$ and $(I,J) \notin \mathcal{N}$, i.e. $J - c_{\mathcal{N}}(I) \neq 0$ (we remind that $c_{\mathcal{N}}(I) = \bigcup\{K \subset N; (I,K) \in \mathcal{N}\}$. Then having $((I,J), c_{\mathcal{N}}(I)) \notin \mathbf{p}$ we get $c_{\mathcal{N}}(I) \notin \mathcal{N}^*$ and existence of a pair $(K,L) \in \mathcal{N}$ for which $((K,L), c_{\mathcal{N}}(I)) \notin \mathbf{p}$ follows. Hence, $(K,L) \in \mathcal{N}, K \subset c_{\mathcal{N}}(I), L - c_{\mathcal{N}}(I) \neq \emptyset, (I, c_{\mathcal{N}}(I)) \in \mathcal{N}$ and using the properties of the FD-relation $\mathcal{N}$ we obtain $(I, L \cup c_{\mathcal{N}}(I))$ what contradicts the definition of $c_{\mathcal{N}}$.

Similarly, the equality $\mathcal{C}^{**} = \mathcal{C}$ for any closure system $\mathcal{C}$ on N take place. Only the nontrivial inclusion $\subset$ will occupy our attention. For $M \in \mathcal{C}^{**} - \mathcal{C}$ let L be the closure of M with respect to $\mathcal{C}$, i.e. $L = \bigcap\{K \in \mathcal{C}; K \supset M\}$, and $L \in \mathcal{C}, L \neq M$. Then

$$((M, L - M), M) \notin \mathbf{p} \Rightarrow (M, L - M) \notin \mathcal{C}^* \Rightarrow \exists K \in \mathcal{C} : ((M, L - M), K) \notin \mathbf{p}$$

and we see that $K \supset M, L - (M \cup K) \neq \emptyset$ and $K \cap L \supset M, K \cap L \in \mathcal{C}$. Thus, we arrived at contradiction with the definition of L.

The proof of first assertion of Theorem is just being over. Namely, from $\mathcal{M} \subset \mu(\mathcal{M})$ we obtain $\mathcal{M}^{**} \subset (\mu(\mathcal{M}))^{**} = \mu(\mathcal{M}) \subset \mathcal{M}^{**}$ and for the closure operators $\mathcal{D} \to \mathcal{D}^{**}$ and δ analogically. The second part of Theorem is a consequence of the general properties of every Galois connection.

$\square$

At the end we comment Theorem. When comparing it with the previous results it may be quoted that it comprehends in a concise and symmetric form Theorems 3, 4 and 7 of [1] and also some results of [6] (e.g. in our setting First Delobel-Casey theorem can be stated in the form

$$\mu(\mathcal{M}) = \mu(\mathcal{N}) \Leftrightarrow \mathcal{M}^{**} = \mathcal{N}^{**}, \qquad \mathcal{M}, \mathcal{N} \in \mathcal{Q}(N)).$$

Besides these facts, and new partial assertions contained there, the language of Galois connections removes a substantial part of proofs, gives a natural dual description of FD-relations and can be succesfully used to an investigation of other dependence or independence structures (see [13]).

ACKNOWLEDGEMENT

The author is grateful to Mr. I. Kramosil for many valuable comments and careful scrutiny of the manuscript.

REFERENCES

[**1**] W. W. Armstrong, Dependency structures of database relationships, In: *Proc. Information Processing 74* (North Holland, Amsterdam, 1974) 580-583.

[**2**] C. Beeri, On the membership problem for functional and multivalued dependencies in relational databases, *ACM Trans. Database Syst.* 3 (Sept. 1980) 241-259.

[**3**] C. Beeri, M. Dowd, R. Fagin and R. Statman, On the structure of Armstrong relations for functional dependencies, *Journal of ACM* 1 (January 1984) 30-46.

[**4**] G. Birkhoff, *Lattice Theory* (Providence R. I., 1967) (Russian translation: Nauka, Moscow, 1984).

[**5**] P. M. Cohn, *Universal Algebra* (D. Reidel Publishing Company, Dordrecht - Boston - London, 1981).

[6] C. Delobel and R. G. Casey, Decomposition of a database and the theory of boolean switching functions, *IBM J. Res. and Develop.* 5 (Sept. 1973) 374-386.

[7] V. Dlab, General algebraic dependence structures and some applications, *Colloquium Mathematicum* 14 (1966) 265-273.

[8] R. Fagin, Functional dependency in a relational database and propositional logic, *IBM J. Res. and Develop.* 6 (Nov. 1977) 534-544.

[9] G. Grätzer, *General Lattice Theory* (Akademie-Verlag, Berlin, 1978).

[10] F. M. Malvestuto, Theory of random observables in relational data bases, *Inform. Systems* 4 (1983) 281-289.

[11] E. Marczewski, A general scheme of the notions of independence in mathematics, *Bulletin de l'Académie Polonaise des Sciences, Série des sci. math., astr. et phys.* 6 (1958) 731-736

[12] N. F. G. Martin and J. W. England, *Mathematical Theory of Entropy* (Addison-Wesley Publ. Comp., London - Amsterdam - Don Mills - Ontario - Sydney - Tokyo, 1981).

[13] F. Matúš, Independence and Radon projections on compact groups, Ph.D. Thesis, Institute of Inf. Th. and Aut., Prague, 1989.

[14] A. W. Naylor, On decomposition theory: generalized dependence, *IEEE Trans. on SMC* 10 (Oct. 1981) 699-713.

[15] D. Geiger and J. Pearl, Logical and algorithmic properties of conditional independence, to appear in: *Proc. 2nd International Workshop on Art. Intel. and Stat.* (Fort Lauderdale, Florida, 1989).

[16] P. C. Saxena and R. C. Tripathi, Cancellation rule and a complete axiomatization for the set of functional dependencies, *Comp. and Art. Intel.* 4 (1989) 347-356.

[17] P. C. Saxena and R. C. Tripathi, An algorithm to find minimal cover of a set of functional dependencies, *Comp. and Art. Intel.* 3 (1988) 277-285.

[18] J. Neveu, *Bases mathématique du Calcul des Probabilités* (Mason et C[ie], Paris, 1964).

[19] J. D. Ullman, *Principles of Database Systems* (Comput. Sci. Press, Potomac, Md., 1982).

[20] N. White, ed., *Theory of Matroids* (Cambridge University Press, London - New York - New Rochelle - Melbourne - Sydney, 1986).

0	1	1
1	2	2
2	7	5
3	61	19
4	2480	184

Table Numbers of all FD-relations and their types on an n-element set, $0 \leq n \leq 4$.