

# Classes of Four-Fold Table Quantifiers

Jan Rauch

Laboratory of Intelligent Systems, Faculty of Informatics and Statistics, University of Economics, W. Churchill Sq. 4, 13067 Prague, Czech Republic rauch@vse.cz

**Abstract.** Four-fold table logical calculi are defined. Formulae of these calculi correspond to patterns based on four-fold contingency tables of two Boolean attributes. An FFT quantifier is a part of the formula, it corresponds to an assertion concerning frequencies from four-fold table. Several classes of FFT quantifiers are defined and studied. It is shown that each particular class has interesting properties from the point of view of KDD. Deduction rules concerning formulae of four-fold tables calculi are demonstrated. It is shown that complex computation of statistical tests can be avoided by using tables of critical frequencies.

## 1 Introduction

Interesting patterns - assertions concerning analyzed database are the the core of KDD process . Some of these assertions can be easily understood as formulae of a suitable logical calculus. Names of relations and names of fields of the analyzed database belong to basic symbols of such calculus. There are interesting and useful features of these calculi, e.g., deduction rules [7].

There is a group of important patterns concerning two Boolean attributes derived from the analyzed database. The patterns correspond to the relations of the Boolean attributes. The patterns are evaluated on the basis of a four-fold table Tab.1. Here  $\varphi$  and  $\psi$  are attributes,  $a$  is the number of objects (records of the analyzed database) satisfying both  $\varphi$  and  $\psi$ ,  $b$  is the number of objects satisfying  $\varphi$  and not satisfying  $\psi$ , etc.

	$\psi$	$\neg\psi$
$\varphi$	$a$	$b$
$\neg\varphi$	$c$	$d$

**Table 1.** Four-fold table of  $\varphi$  and  $\psi$

An example of a pattern based on the four-fold table is the *association rule*, see [1]. Attributes  $\varphi$  and  $\psi$  with four-fold table  $\langle a, b, c, d \rangle$  are associated by an association rule with parameters *conf* and *sup* if and only if

$$\frac{a}{a+b} \geq \text{conf} \wedge \frac{a}{a+b+c+d} \geq \text{sup} .$$

Further examples are given, e.g., in [2], [8], [9].

Observational calculi are logical calculi formulae of which correspond to various assertions concerning analyzed data. They are defined and studied in [2], see also [7]. A special case of observational calculi is defined in section 2. It is a *Four-Fold Table Predicate Calculus*, further only FFTPC. The formulae of FFTPC are of the form  $\varphi \sim \psi$ , where  $\varphi$  and  $\psi$  are Boolean attributes. Symbol  $\sim$  is an *FFT quantifier*. It expresses the relation of attributes  $\varphi$  and  $\psi$ .

The goal of this paper is to show that it is useful to understand the relations of two Boolean attributes as formulae of FFTPC. Several classes of FFT-quantifiers are defined in section 3, e.g., implicational or equivalence quantifiers. Deduction rules concerning the formulae of FFTPC are studied in section 4. Tables of critical frequencies are discussed in section 5. They can be used to avoid a complex computation when testing patterns of the form  $\varphi \sim \psi$ .

## 2 Four-fold Table Predicate Calculi

We have to define a language of FFTPC, its models and values of formulae.

**Definition 1.** The **type of FFTPC** is an integer positive number  $T$ .

**Definition 2.** A **language of the FFTPC of type  $T$**  is given by:

1. **Basic symbols:**

*Basic attributes*  $A_1, \dots, A_T$ , *propositional connectives*  $\wedge, \vee, \neg$ , and *FFT quantifier*  $\sim$ .

2. **Attributes:**

- If  $\varphi$  is a basic attribute, then  $\varphi$  is an attribute.
- If  $\varphi$  and  $\psi$  are attributes, then also  $\varphi \wedge \psi$ ,  $\varphi \vee \psi$  and  $\neg\varphi$  are derived attributes.
- If  $\varphi$  is a derived attribute, then  $\varphi$  is an attribute.
- Usual conventions concerning parentheses are valid.

3. **Formulae:**

If  $\varphi$  and  $\psi$  are attributes and  $\sim$  is an FFT quantifier, then  $\varphi \sim \psi$  is a formula.

**Definition 3.** A **model of the FFTPC of type  $T$**  is each  $\{0,1\}$ -data matrix with  $T$  columns.

**Definition 4.** We consider each model  $M$  with  $n$  rows to be the result of an observation of  $n$  objects. The  $i$ -th observed object corresponds to the  $i$ -th row. We say that the  $i$ -th **object has the basic attribute**  $A_1$  in  $M$  if value 1 is in the first column of the  $i$ -th row of  $M$ . We say that the  $i$ -th object has derived attribute  $A_1 \wedge A_2$  in  $M$  if value 1 is both in the first and in the second column of the  $i$ -th row of  $M$ . Similarly for other basic and derived attributes.

**Definition 5.** **Frequency**  $Fr(\varphi, M)$  **of an attribute**  $\varphi$  **in a model**  $M$  is the number of objects having attribute  $\varphi$ .

**Definition 6. Associated function  $F_{\sim}$  of the FFT quantifier  $\sim$**  is a  $\{0, 1\}$ -valued function defined for all quadruples  $\langle a, b, c, d \rangle$  of non-negative integer numbers such that  $a + b + c + d > 0$ . We usually write only  $\sim(a, b, c, d)$  instead of  $F_{\sim}(a, b, c, d)$ .

**Definition 7.** Let  $\varphi \sim \psi$  be a formula of an FFTPC  $\Phi$ . Let  $M$  be a model  $\Phi$ . Then **value**  $Val(\varphi \sim \psi, M)$  **of the formula  $\varphi \sim \psi$  in the model  $M$**  is defined as the value

$$F_{\sim}(Fr(\varphi \wedge \psi, M), Fr(\varphi \wedge \neg\psi, M), Fr(\neg\varphi \wedge \psi, M), Fr(\neg\varphi \wedge \neg\psi, M)) .$$

If  $Val(\varphi \sim \psi, M) = 1$ , then we say that  $\varphi \sim \psi$  is **true in  $M$** . If  $Val(\varphi \sim \psi, M) = 0$ , then  $\varphi \sim \psi$  is **false in  $M$** .

The quadruple  $\langle a, b, c, d \rangle$ , where  $a = Fr(\varphi \wedge \psi, M)$ ,  $b = Fr(\varphi \wedge \neg\psi, M)$ ,  $c = Fr(\neg\varphi \wedge \psi, M)$  and  $d = Fr(\neg\varphi \wedge \neg\psi, M)$ , is called the four-fold table of the formula  $\varphi \sim \psi$  in a model (data matrix)  $M$ . We shall write only FFT instead of four-fold table.

### 3 Classes of Quantifiers

Properties of a formula  $\varphi \sim \psi$  depend on properties of its associated function  $F_{\sim}(a, b, c, d)$ . We define the function  $F_{\sim}$  to make the pattern  $\varphi \sim \psi$  interesting from the point of view of a database user. One of the interesting relations of attributes  $\varphi$  and  $\psi$  is the relation  $\varphi$  *implies*  $\psi$ . An important question is: "Which quantifier expresses the relation of implication?" A class of implicational quantifiers is defined in [2].

**Definition 8.** An FFT quantifier  $\sim$  is **implicational** if

$$\sim(a, b, c, d) = 1 \wedge a' \geq a \wedge b' \leq b \text{ implies } \sim(a', b', c', d') = 1$$

for non-negative integers  $a, b, c, d, a', b', c', d'$  such that  $a + b + c + d > 0$  and  $a' + b' + c' + d' > 0$ .

The condition  $a' \geq a$  and  $b' \leq b$  means that the four-fold table  $\langle a', b', c', d' \rangle$  is "better from the point of view of implication" than the four-fold table  $\langle a, b, c, d \rangle$  (i-better, see [2]). If  $\langle a, b, c, d \rangle$  is an FFT of  $\varphi \sim \psi$  in model  $M$  and  $\langle a', b', c', d' \rangle$  is an FFT of  $\varphi \sim \psi$  in model  $M'$ , then the sentence "Better from the point of view of implication" means: In model  $M'$  are more objects satisfying both  $\varphi$  and  $\psi$  than in  $M$ , and in model  $M'$  are fewer objects satisfying  $\varphi$  and not satisfying  $\psi$  than in  $M$ . Thus it is reasonable to expect that if formula  $\varphi \sim \psi$  is true in model  $M$ , then it is also true in model  $M'$ . This expectation is ensured for implicational quantifiers by definition 8.

*Example 1.* Quantifier  $\Rightarrow_{p,s}$  of *founded implication* for  $0 < p \leq 1$  and  $s > 0$  is implicational, see [2]. It is  $\Rightarrow_{p,s}(a, b, c, d) = 1$  if and only if  $\frac{a}{a+b} \geq p \wedge a \geq s$ .

*Example 2.* Quantifier  $\Rightarrow_{p,\alpha,s}^!$  of lower critical implication for  $0 < p < 1$ ,  $0 < \alpha < 1$  and  $s > 0$  is implicational, see [2]. It is

$$\Rightarrow_{p,\alpha,s}^! (a, b, c, d) = 1 \text{ if and only if } \sum_{i=a}^{a+b} \frac{(a+b)!}{i!(a+b-i)!} p^i (1-p)^{a+b-i} \leq \alpha \wedge a \geq s .$$

Let us note that the formula  $\varphi \Rightarrow_{p,\alpha,s}^! \psi$  corresponds to the test (on the level  $\alpha$ ) of the null hypothesis  $H_0 : P(\psi|\varphi) \leq p$  against the alternative one  $H_1 : P(\psi|\varphi) > p$ . Here  $P(\psi|\varphi)$  is the conditional probability of the validity of  $\psi$  under the condition  $\varphi$ , see [2] for more details.

It is easy to prove for an implicational quantifier  $\Rightarrow^*$  that the value  $\Rightarrow^*(a, b, c, d)$  does not depend either on  $c$  or on  $d$ . Thus we shall write only  $\Rightarrow^*(a, b)$  instead of  $\Rightarrow^*(a, b, c, d)$  for each implicational quantifier  $\Rightarrow^*$ .

We call conditions like " $a' \geq a$  and  $b' \leq b$ " **truth preservation conditions**, see [8]. The condition  $a' \geq a$  and  $b' \leq b$  is a **truth preservation condition for implicational quantifiers**. Further classes of FFT quantifiers are defined in [8], some of them using truth preservation conditions.

**Definition 9.** An FFT quantifier  $\sim$  is double implicational if

$$\sim (a, b, c, d) = 1 \wedge a' \geq a \wedge b' \leq b \wedge c' \leq c \text{ implies } \sim (a', b', c', d') = 1$$

for non-negative integers  $a, b, c, d, a', b', c', d'$  such that  $a + b + c + d > 0$  and  $a' + b' + c' + d' > 0$ .

We can see a reason for such a definition in an analogy to propositional logic. If  $u$  and  $v$  are propositions and both  $u \rightarrow v$  and  $v \rightarrow u$  are true, then  $u$  is equivalent to  $v$  ( $\rightarrow$  is a propositional connective of implication). Thus we can try to express the relation of equivalence of attributes  $\varphi$  and  $\psi$  using "double implicational" FFT-quantifier  $\Leftrightarrow^*$  such that

$$\varphi \Leftrightarrow^* \psi \text{ if and only if } \varphi \Rightarrow^* \psi \wedge \psi \Rightarrow^* \varphi,$$

where  $\Rightarrow^*$  is a suitable implicational quantifier. If we apply the truth preservation condition for implicational quantifier to  $\varphi \Rightarrow^* \psi$ , we obtain  $a' \geq a \wedge b' \leq b$ . If we apply it to  $\psi \Rightarrow^* \varphi$ , we obtain  $a' \geq a \wedge c' \leq c$ , ( $c$  is here instead of  $b$ , see Tab.1). This leads to **the truth preservation condition for double implicational quantifiers**  $a' \geq a \wedge b' \leq b \wedge c' \leq c$ , see definition 9. Several FFT quantifiers are defined according to this idea in [3], an example follows.

*Example 3.* A quantifier  $\Leftrightarrow_{p,s}$  of founded double implication for  $0 < p \leq 1$  and  $s > 0$  is double implicational, see [8]. It is  $\Leftrightarrow_{p,s} (a, b, c, d) = 1$  if and only if  $\frac{a}{a+b+c} \geq p \wedge a \geq s$ .

It is proved in [8] that quantifier  $\Leftrightarrow_{p,s}$  belongs to the class of  $\Sigma$ -double implicational quantifiers:

**Definition 10.** An FFT quantifier  $\sim$  is  $\Sigma$ -double implicational if

$$\sim(a, b, c, d) = 1 \wedge a' \geq a \wedge b' + c' \leq b + c \text{ implies } \sim(a', b', c', d') = 1$$

for non-negative integers  $a, b, c, d, a', b', c', d'$  such that  $a + b + c + d > 0$  and  $a' + b' + c' + d' > 0$ .

It is obvious that each  $\Sigma$ -double implicational quantifier is also double implicational. It follows from the definition that if a quantifier  $\Leftrightarrow^*$  belongs to  $\Sigma$ -double implicational quantifiers, then the value  $\Rightarrow^*(a, b, c, d)$  does not depend on  $d$ . Thus we shall write only  $\Leftrightarrow^*(a, b, c)$  instead of  $\Leftrightarrow^*(a, b, c, d)$  for  $\Sigma$ -double implicational quantifier  $\Leftrightarrow^*$ .

We have a similar situation for equivalence quantifiers. If  $u$  and  $v$  are propositions and both  $u \rightarrow v$  and  $\neg u \rightarrow \neg v$  are true, then  $u$  is equivalent to  $v$ . Thus we can try to express the relation of equivalence of attributes  $\varphi$  and  $\psi$  using an "equivalence" FFT-quantifier  $\equiv^*$  such that

$$\varphi \equiv^* \psi \text{ if and only if } \varphi \Rightarrow^* \psi \wedge \neg\varphi \Rightarrow^* \neg\psi,$$

where  $\Rightarrow^*$  is a suitable implicational quantifier. If we apply the truth preservation condition for implicational quantifier to  $\varphi \Rightarrow^* \psi$  we obtain  $a' \geq a \wedge b' \leq b$ , if we apply it to  $\neg\varphi \Rightarrow^* \neg\psi$ , we obtain  $d' \geq d \wedge c' \leq c$ , ( $c$  is here instead of  $b$  and  $d$  instead of  $a$ , see Tab.1). This leads to the *truth preservation condition for equivalence quantifiers*:  $a' \geq a \wedge b' \leq b \wedge c' \leq c \wedge d' \geq d$  and consequently to the definition of **equivalence quantifiers** [8] (associational quantifiers according to [2]). In [3] are defined several FFT quantifiers as equivalence quantifiers.

*Example 4.* A quantifier  $\equiv_p$  of  $p$ -equivalence for  $0 < p \leq 1$  is equivalence quantifier, see [8]. It is  $\Leftrightarrow_{p,s}(a, b, c, d) = 1$  if and only if  $\frac{a}{a+b+c} \geq p \wedge a \geq s$ . The quantifier  $\equiv_p$  is also a  $\Sigma$ -equivalence quantifier [8].

Several further classes of implicational quantifiers are defined and studied in [8], e.g., *pure double implicational, typical double implicational, pure equivalence, and typical  $\Sigma$ -equivalence*. Each of these classes of FFT quantifiers contains useful quantifiers. Some of them are implemented in the system PC-GUHA, see [6]. Some important features of FFT-quantifiers related to classes of FFT are discussed in the following sections.

## 4 Deduction rules

A deduction rule is the relation of the form

$$\frac{\alpha_1, \dots, \alpha_n}{\beta},$$

where  $\alpha_1, \dots, \alpha_n, \beta$  are formulae. This **deduction rule is correct** if for each model  $M$  holds: if  $\alpha_1, \dots, \alpha_n$  are true in  $M$ , then also  $\beta$  is true in  $M$ . We are interested in correct deduction rules of the the form  $\frac{\varphi \sim \psi}{\varphi' \sim \psi'}$ , where  $\varphi \sim \psi$  and  $\varphi' \sim \psi'$  are formulae of an FFTPC. Such deduction rules can be used, e.g., in the following ways:

- **To reduce the output of a data mining procedure:** If formula  $\varphi \sim \psi$  is a part of a data mining procedure output (thus it is true in analyzed data) and if  $\frac{\varphi \sim \psi}{\varphi' \sim \psi'}$  is a correct deduction rule, then it is not necessary to put  $\varphi' \sim \psi'$  into the output. The used deduction rule must be clear enough from the point of view of the user of the data mining procedure. An example of a simple deduction rule is dedereduction deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi \Rightarrow^* \psi \vee \chi}$ , which is correct for each implicational quantifier  $\Rightarrow^*$  [2].
- **To decrease the number of actually tested formulae:** If formula  $\varphi \sim \psi$  is true in an analyzed model (data matrix) and if  $\frac{\varphi \sim \psi}{\varphi' \sim \psi'}$  is a correct deduction rule, then it is not necessary to test  $\varphi' \sim \psi'$ .

Let us note that correct deduction rules (not only in the form of  $\frac{\varphi \sim \psi}{\varphi' \sim \psi'}$ ) are used in the GUHA procedure PC-ASSOC, see, e.g., [4] or [6]. Anyway, it is reasonable to ask when the deduction rule of the form  $\frac{\varphi \sim \psi}{\varphi' \sim \psi'}$  is correct. Several results concerning this problem were achieved in [5]. We are going to show some of them. We need several notions.

**Definition 11.** Let  $\varphi$  be an attribute. Then an **associated propositional formula**  $\pi(\varphi)$  to  $\varphi$  is the same string of symbols as  $\varphi$ , but the particular basic attributes are understood as the propositional variables.

*Example 5.* If  $A_1 \wedge A_7$  is a derived attribute, then  $\pi(A_1 \wedge A_7)$  is propositional formula  $A_1 \wedge A_7$  with propositional variables  $A_1$  and  $A_7$ .

**Definition 12.** Let  $\sim$  be an FFT quantifier. Then

1.  $\sim$  is **a-dependent** if there are non-negative integers  $a, a', b, c, d$  such that  $\sim(a, b, c, d) \neq \sim(a', b, c, d)$ . Analogously for **b-dependent**.
2.  $\sim$  is **(b+c)-dependent** if there are non-negative integers  $a, b, c, d, b', c'$  such that  $b + c \neq b' + c'$  and  $\sim(a, b, c, d) \neq \sim(a, b', c', d)$ . Analogously for **(a+d)-dependent**.

**Definition 13.** Interesting quantifiers:

1. An **implicational quantifier  $\Rightarrow^*$  is interesting** if  $\Rightarrow^*$  is both a-dependent and b-dependent, and if  $\Rightarrow^*(0, 0) = 0$ .
2. A  **$\Sigma$ -double implicational quantifier  $\Leftrightarrow^*$  is interesting** if  $\Leftrightarrow^*$  is both a-dependent and (b+c)-dependent, and if  $\Leftrightarrow^*(0, 0, 0) = 0$ .
3. A  **$\Sigma$ -equivalence quantifier  $\equiv^*$  is interesting** if  $\equiv^*$  is (a+d)-dependent and if  $\equiv^*(0, b, c, 0) = 0$  for  $b + c > 0$ .

**Theorem 14.** *Correct deduction rules:*

1. *If  $\Rightarrow^*$  is an interesting implicational quantifier, then deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi' \Rightarrow^* \psi'}$  is correct if and only if at least one of the following conditions a), b) is satisfied (see also [7]):*
  - a) *Both (i) and (ii) are tautologies:*

- (i)  $\pi(\varphi) \wedge \pi(\psi) \rightarrow \pi(\varphi') \wedge \pi(\psi')$ ,  
(ii)  $\pi(\varphi') \wedge \neg\pi(\psi') \rightarrow \pi(\varphi) \wedge \neg\pi(\psi)$ .
- b)  $\pi(\varphi) \rightarrow \neg\pi(\psi)$  is a tautology.
2. If  $\Rightarrow^*$  is an interesting  $\Sigma$ -double implicational quantifier, then deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi' \Rightarrow^* \psi'}$  is correct if and only if at least one of the following conditions a), b) is satisfied:
- a) Both (i) and (ii) are tautologies:  
(i)  $\pi(\varphi) \wedge \pi(\psi) \rightarrow \pi(\varphi') \wedge \pi(\psi')$ ,  
(ii)  $\pi(\varphi') \wedge \neg\pi(\psi') \vee \neg\pi(\varphi') \wedge \pi(\psi') \rightarrow \pi(\varphi) \wedge \neg\pi(\psi) \vee \neg\pi(\varphi) \wedge \pi(\psi)$ .
- b) (i):  $\pi(\varphi) \rightarrow \neg\pi(\psi)$  or (ii):  $\pi(\psi) \rightarrow \neg\pi(\varphi)$  are tautologies.
3. If  $\equiv^*$  is an interesting  $\Sigma$ -equivalence quantifier, then deduction rule  $\frac{\varphi \equiv^* \psi}{\varphi' \equiv^* \psi'}$  is correct if and only if  $(\pi(\varphi) \wedge \pi(\psi) \vee \neg\pi(\varphi) \wedge \neg\pi(\psi)) \rightarrow (\pi(\varphi') \wedge \pi(\psi') \vee \neg\pi(\varphi') \wedge \neg\pi(\psi'))$  is a tautology.

*Proof.* Let us outline the proof of a theorem concerning correct deduction rules for implicational quantifiers, see point 1. We have to prove:

- A):** If 1.a is satisfied, then deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi' \Rightarrow^* \psi'}$  is correct.  
**B):** If 1.b is satisfied, then deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi' \Rightarrow^* \psi'}$  is correct.  
**C):** If neither 1.a nor 1.b are satisfied, then deduction rule  $\frac{\varphi \Rightarrow^* \psi}{\varphi' \Rightarrow^* \psi'}$  is not correct.

We suppose:  $M$  is a model,  $\varphi \Rightarrow^* \psi$  is a formula, and  $\langle a, b, c, d \rangle$  is a contingency table of  $\varphi \Rightarrow^* \psi$  in  $M$ , analogously for  $\varphi' \Rightarrow^* \psi'$  and  $\langle a', b', c', d' \rangle$ .

**A):** Let  $Val(\varphi \Rightarrow^* \psi, M) = 1$ . It means  $\Rightarrow^*(a, b) = 1$ . The condition 1.a(i) implies  $a \leq a'$ , the condition 1.a(ii) implies  $b' \leq b$ . Thus it is also  $\Rightarrow^*(a', b') = 1$ , see the definition 8 of the implicational quantifier.

**B):** If  $\pi(\varphi) \rightarrow \neg\pi(\psi)$  is a tautology, then it is  $a = 0$  for each model  $M$ . The quantifier  $\Rightarrow^*$  is an interesting implicational quantifier, thus  $\Rightarrow^*(0, 0) = 0$ . It implies  $\Rightarrow^*(0, b) = 0$  for each  $b$  (see definition 13 of the interesting implicational quantifier). Thus  $Val(\varphi \Rightarrow^* \psi, M) = 0$  for each model  $M$ . It means that the assertion: *If  $Val(\varphi \Rightarrow^* \psi, M) = 1$  then  $Val(\varphi' \Rightarrow^* \psi', M) = 1$*  is true.

**C):** We suppose that neither 1.a nor 1.b are satisfied. We have to find a model  $M$  such that  $Val(\varphi \Rightarrow^* \psi, M) = 1$  and  $Val(\varphi' \Rightarrow^* \psi', M) = 0$ . If neither 1.a nor 1.b are satisfied, then **D)** or **E)** are satisfied:

- D):** Neither 1.a(i) nor 1.b are satisfied.  
**E):** Neither 1.a(ii) nor 1.b are satisfied.

**D):** The implicational quantifier  $\Rightarrow^*$  is  $a$ -dependent, thus there are  $A, B$  such that  $\Rightarrow^*(A, B) = 1$ . The condition 1.a(i) is not satisfied, thus there is an object  $o$ , such that  $o$  has the attribute  $\varphi \wedge \psi$ , and  $o$  has not the attribute  $\varphi' \wedge \psi'$ . Let  $M$  be a model with  $A$  objects  $o$  (a data matrix with  $A$  identical rows, each of them equal to a row corresponding to the row of  $o$ ). It means  $a = A, b = 0$  and also  $a' = 0, b' \geq 0$ . The quantifier  $\Rightarrow^*$  is implicational and  $\Rightarrow^*(A, B) = 1$ , thus also  $\Rightarrow^*(A, 0) = 1$ . Further,  $a' = 0$  and  $\Rightarrow^*(0, 0) = 0$  ( $\Rightarrow^*$  is interesting

implicational), thus also  $\Rightarrow^* (a', b') = 0$ . It means that  $Val(\varphi \Rightarrow^* \psi, M) = 1$  and  $Val(\varphi' \Rightarrow^* \psi', M) = 0$ .

E): The quantifier  $\Rightarrow^*$  is *b-dependent*, thus there are non-negative integers  $A, B, B'$  such that  $\Rightarrow^* (A, B) = 1$  and  $\Rightarrow^* (A, B') = 0$ . The condition 1.a(ii) is not satisfied, thus there is an object which has the attribute  $\varphi' \wedge \neg\psi'$  and has not the attribute  $\varphi \wedge \neg\psi$ . Let us call it *object of type B*. The condition 1.b is not satisfied, thus there is an object which has both the attribute  $\varphi$  and the attribute  $\psi$ . Let us call it *object of type A*. Let  $M$  be a model with  $A$  objects of type **A** and  $B$  objects of type **B**. Values of attributes important for computing  $Val(\varphi \Rightarrow^* \psi, M)$  and  $Val(\varphi' \Rightarrow^* \psi', M)$  and consequences for contingency tables  $\langle a, b, c, d \rangle$  and  $\langle a', b', c', d' \rangle$  are in Tab. 2. Symbol " - " means that we do not know

line	attribute	A objects of type <b>A</b>	B objects of type <b>B</b>	consequence
1	$\varphi \wedge \psi$	<b>X</b> : $T$ by definition	-	$a \geq A$
2	$\varphi' \wedge \psi'$	-	$B$ , see <b>Y</b>	$a' \leq A$
3	$\varphi \wedge \neg\psi \vee \neg\varphi \wedge \psi$	$F$ , see <b>X</b>	$F$ by definition	$b = 0$
4	$\varphi' \wedge \neg\psi' \vee \neg\varphi' \wedge \psi'$	-	<b>Y</b> : $T$ by definition	$b' \geq B$

Table 2. Values of attributes for case E

the corresponding value, **X**: and **Y**: are labels. Symbol  $T$  means that an object of a corresponding type has a corresponding attribute (symbol  $F$  means "has not"). It is  $\Rightarrow^* (A, B) = 1$ ,  $a \geq A$  and  $b = 0$ , thus also  $\Rightarrow^* (a, b) = 1$ . Further, it is  $\Rightarrow^* (A, B') = 0$ ,  $a' \leq A$  and  $b' \geq B'$ , thus  $\Rightarrow^* (a, b') = 0$ . This finishes the proof.  $\square$

## 5 Tables of critical frequencies

**Theorem 15.** *Let  $\Rightarrow^*$  be an implicational quantifier. Then there is a non-negative and non-decreasing function  $Tb_{\Rightarrow^*}$  with value  $Tb_{\Rightarrow^*}(a) \in \{0, 1, 2, \dots\} \cup \{\infty\}$  such that it is*

$$\Rightarrow^* (a, b) = 1 \text{ if and only if } b < Tb_{\Rightarrow^*}(a)$$

for all integers  $a \geq 0$  and  $b \geq 0$ .

*Proof.* We define  $Tb_{\Rightarrow^*}(a) = \min\{e \mid \Rightarrow^* (a, e) = 0\}$ , see the definition of implicational quantifiers.  $\square$

We call function  $Tb_{\Rightarrow^*}$  a *table of critical frequencies for implicational quantifier  $\Rightarrow^*$* . It is used in the GUHA procedure PC-ASSOC [7]. It is important that the function  $Tb_{\Rightarrow^*}$  makes it possible to use a simple test of inequality instead of a rather complex computation. E.g., we can use inequality  $b < Tb_{\Rightarrow^*}^!_{p, \alpha, s}(a)$  instead of condition  $\sum_{i=a}^{a+b} \frac{(a+b)!}{i!(a+b-i)!} p^i (1-p)^{a+b-i} \leq \alpha \wedge a \geq s$  for quantifier  $\Rightarrow^!_{p, \alpha, s}$  of

lower critical implication, see example 2. An other form of the table of critical frequencies for implicational quantifier is defined in [2]. Further tables of critical frequencies for  $\Sigma$ -double implicational quantifiers and for  $\Sigma$ -equivalence quantifiers are defined in [8].

## 6 Conclusions

We have defined several classes of FFT quantifiers. Deduction rules and tables of critical frequencies have been discussed as useful tools for dealing with some patterns in the KDD process. We have shown that properties of these tools are closely related to classes of FFT quantifiers.

There are further useful classes of FFT quantifiers, e.g., symmetrical quantifiers, a,d-symmetrical quantifiers, strong double implicational quantifiers, strong double equivalence quantifiers and F-quantifiers (with the same behaviour as Fisher's test), see [2], [5], [8].

*This work is supported by grant 47160008 of the Ministry of Education and by grant 201/96/1445 of the Grant Agency of the Czech Republic.*

## References

1. Aggraval, R. et al: Fast Discovery of Association Rules. In Fayyad, U. M. et al.: Advances in Knowledge Discovery and Data Mining. AAAI Press / The MIT Press, 1996. 307–328
2. Hájek, P., Havránek T.: Mechanising Hypothesis Formation - Mathematical Foundations for a General Theory. Berlin - Heidelberg - New York, Springer-Verlag, 1978, 396 p.
3. Hájek, P. - Havránek, T., Chytil M.: Metoda GUHA. Praha, Academia, 1983, 314 p. (in Czech)
4. Hájek, P., Sochorová, A., Zvárová, J.: GUHA for personal computers. Computational Statistics & Data Analysis **19**, (1995) 149–153
5. Rauch, J.: Logical foundations of mechanizing hypotheses formation from databases (in Czech). Thesis, Mathematical Institute of Czechoslovak Academy of Sciences Prague, 1986, 133 p.
6. Rauch, J.: GUHA as a Data Mining Tool. In: Practical Aspects of Knowledge Management. Schweizer Informatiker Gesellschaft Basel, 1996
7. Rauch, J.: Logical Calculi for Knowledge Discovery in Databases. In Principles of Data Mining and Knowledge Discovery, (J. Komorowski and J. Zytkow, eds.), Springer Verlag, Berlin, 47-57, 1997.
8. Rauch, J.: Four-Fold Table Calculi. LISp, Technical Report LiSp9710 (in Czech), 1998
9. Zembowicz R. - Zytkow J.: From Contingency Tables to Various Forms of Knowledge in Databases. In Fayyad, U. M. et al.: Advances in Knowledge Discovery and Data Mining. AAAI Press/ The MIT Press, 1996. s. 329 - 349.