THE LOGIC OF PREDICTION

EVGENII VITYAEV
Sobolev Institute of Mathematics, Russian Academy of Science,
Koptyug prospect 4, Novosibirsk, 630090, Russia,
E-mail: vityaev@math.nsc.ru

We consider the predictions provided by the inductive theories. For these theories predictions are performed by the Inductive Statistical (I-S) inferences. It was noted by Hempel that the I-S inference is statistically ambiguous. To avoid this ambiguity we need to use the rules that satisfy the Requirement of Maximum Specificity (RMS). The formal definition of the RMS wasn't given by Hempel. We define the notions of law and probabilistic law, and also the sets of all laws $\mathcal{L}$, and probabilistic laws $\mathcal{LP}$. We prove that the set SPL of Strongest Probabilistic Laws (with the maximum values of conditional probability) contains the set $\mathcal{L}$, so we have $\mathcal{L} \subset \text{SPL} \subset \mathcal{LP}$. We prove that the maximum specific rules - the strongest SPL rules for prediction of atoms - satisfy the RMS condition. The maximum specific rules may be used in I-S inference. We prove that the set MSR of all Maximum Specific Rules is consistent and the I-S inferences based on MSR rules avoid the problem of statistical ambiguity. We define Semantic Probabilistic Inferences (SP-inference) that infer the sets $\mathcal{L}, \mathcal{LP}$, SPL, MSR. Finally, we mention the program system 'Discovery', which realize the SP-inference and discovers the sets $\mathcal{L}, \mathcal{LP}$, SPL, MSR. This system was applied for solution of many practical tasks (see website www.math.nsc.ru/AP/ScientificDiscovery).

1. Induction

1.1. *The statistical ambiguity problem*

One of the major results of the Philosophy of Science is so-called *Covering Law Model* that was introduced by Hempel in the early sixties in his famous article 'Aspects of Scientific Explanation' (see Hempel [1,2], and Salmon [3] for a historical overview). The basic idea of this covering law model is that a fact is explained by subsumption under so-called *covering law,* i.e. the task of an explanation is to show that a fact can be considered as an instantiation of a law. In the covering law model two types of explanation are distinguished: *Deductive-Nomological* explanations (D-N explanations) and *Inductive-Statistical* explanations (I-S explanations). In D-N explanations the law is *deterministic*, whereas in I-S explanations the

law is *statistical.* Right from the beginning it was clear to Hempel that two I-S explanations can yield contradictory conclusions. He called this phenomenon the *statistical ambiguity* of I-S explanations [1,2]. Let us consider the following example of the statistical ambiguity.

Suppose that we have the following statements about Jane Jones. 'Almost all cases of streptococcus infection clear up quickly after the administration of penicillin'(L1). 'Almost no cases of penicillin resistant streptococcus infection clear up quickly after the administration of penicillin'(L2). 'Jane Jones had streptococcus infection'(C1). 'Jane Jones received treatment with penicillin'(C2). 'Jane Jones had a penicillin resistant streptococcus infection'(C3). From these statements it is possible to construct two contradictory arguments, one explaining why Jane Jones recovered quickly (E), and the other one, explaining its negation why Jane Jones did not recover quickly (¬E).

$$\begin{array}{cc} Argument1 & Argument2 \\ \dfrac{\dfrac{L1}{C1, C2}}{E}[r] & \dfrac{\dfrac{L2}{C2, C3}}{\neg E}[r] \end{array}$$

The premises of both arguments are consistent with each other, they could all be true. However, their conclusions contradict each other, making these arguments rival ones.

Hempel hoped to solve this problem by forcing all statistical laws in an argument to be maximally specific. That is, they should contain all relevant information with respect to the domain in question. In our example, then, premise C3 of the second argument invalidates the first argument, since the law L1 is not maximally specific with respect to all information about Jane Jones. So, we can only explain ¬E, but not E.

1.2. *Inductive-Statistical Inference*

Hempel proposed the formalization of the statistical inference as Inductive-Statistical Inference (I-S inference) and the property of the maximal specific statistical laws as the Requirement of Maximal Specificity (RMS). The Inductive-Statistical Inference has the form:

$$\dfrac{\dfrac{L_1, \ldots, L_{\mathrm{m}}}{C_1, \ldots, C_{\mathrm{n}}}}{G}[r]$$

It satisfies the following conditions:

- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$, $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n} \vdash$ G;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$, $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}$ are consistent;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m} \nvdash$ G, $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n} \nvdash$ G;
- $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$ are composed of statistical quantified formulas.
- $\mathrm{C}_1,\ldots,\mathrm{C}_\mathrm{n}$ are quantifier-free;
- RMS: All laws $\mathrm{L}_1,\ldots,\mathrm{L}_\mathrm{m}$ are maximal specific.

In Hempel's [1,2] the RMS is defined as follows. An I-S argument of the form:

$$\frac{\dfrac{p(G;F)}{F(\boldsymbol{a})}}{G(\boldsymbol{a})}[r]$$

is an acceptable I-S explanation with respect to a "knowledge state" K, if the following Requirement of Maximal Specificity is satisfied. For any class H for which the following two sentences are contained in K

$$\begin{aligned} &\forall x(H(x) \Rightarrow F(x)), \\ &H(\boldsymbol{a}), \end{aligned} \tag{1}$$

there exists a statistical law p(G;H) = r' in K such that r = r'. The basic idea of RMS is that if F and H both contain the object $\boldsymbol{a}$, and H is a subset of F, then H provides more specific information about the object $\boldsymbol{a}$ than F, and therefore the law p(G;H) should be preferred over the law p(G; F).

1.3. *The Requirement of Maximal Specificity in default logic*

Nowadays the same problems arise in non-monotonic logic and especially in default logic. Hempel's RMS produces also non-monotonic effects in inductive statistical reasoning. The streptococcus infection example is non-monotonic in the following sense. It was observed that the conflict between argument 1 and the argument 2 depends on the knowledge state K. If K contains only the information that John is infected, then RMS determines that argument 1 is the best explanation. In that case K implies the conclusion that John will recover quickly. However, if K is expanded with the premise C3, i.e. the information that John had a penicillin resistant streptococcus infection, then RMS determines that argument 2 is the best explanation and John will not recover quickly. Hence, the conclusion that John will recover quickly is not preserved under expansion of K.

Yao-Hua Tan [4] showed that there is a remarkable resemblance between two research traditions: default logic and inductive-statistical explanations.

Both research traditions have the same research objective; to develop formalisms for reasoning with incomplete information. In both research traditions the crucial problem that had to be dealt with is the problem of *Specificity*, i.e. when two arguments conflict with each other the most specific argument has to be preferred to the less specific argument. This criterion of specificity that was proposed in AI research is very similar to the criterion of maximal specificity suggested by Hempel in the early sixties.

Let us formulate the Requirement of Maximal Specificity (RMS*) in default logic. Essentially, default logic is an ordinary first-order predicate logic extended with extra inference rules that are called default rules. The logical form of a *default rule* follows:

$$(\alpha(x) : \beta_1(x), \ldots, \beta_n(x)/\omega(x))$$

The subformulas α(x), β_i(x), and ω(x) are predicate logical formulas with free variable x. The subformula α(x) is called the *prerequisite*, β_i(x) are the *justifications* and ω(x) is the *consequent* of the default rule. The intuitive interpretation of a default rule follows: if the prerequisite α(x) is valid, and all justifications β_i(x) are consistent with the available information (i.e. $\neg\beta_i$(x) is not derivable from the available information), then one can assume that the consequent ω(x) is valid.

A set of formulas E is an *extension* of the default theory $\Delta = \langle W;D \rangle$, D – the set of default rules, W – a set of predicate logical formulas, if E is the smallest set such as: $W \subset E$; E = Th(E); for each default rule $(\alpha(x):\beta_1(x),\ldots,\beta_n(x)/\omega(x)) \subset D$, and each term t: if $\alpha(t) \in E$, and $\neg\beta_1(t),\ldots,\neg\beta_n(t) \notin E$, then $\omega(t) \in E$.

RMS*: If a default theory has multiple conflicting extensions, then the extension is preferred which is generated by the most specific defaults [4].

The default rule with the 'most specific' prerequisite is preferred in case of conflicts. Let A(x) and B(x) be the prerequisites of the default rules D1 and D2. The prerequisite A(x) is *more specific* than B(x) if the set that the predicate A refers to is a subset of the set that B refers to, i.e. if the sentence $\forall x(A(x) \Rightarrow B(x))$ is valid. It is obvious that this criterion can be considered as the analogue of RMS in default logic.

1.4. *The solution of the statistical ambiguity problem*

From the previous consideration we see that the statistical ambiguity problem raises in AI in different forms, but it isn't solved hitherto. We will once again state the problem that wasn't solved by Hempel and his followers:

Statistical Ambiguity Problem. Is it possible to define the RMS in such a way that it solves the statistical ambiguity problem? Can we define the RMS in such a way that the set of sentences satisfying the RMS be consistent?

This problem is very important, because it means the consistency of predictions. The predictions nowadays are produced by different AI systems: expert systems, knowledge bases, robotics, intelligent data analysis and etc.

In this paper we present the solution of this problem. We define the set of Maximum Specific Rules (MSR) and the Requirement of Maximal Specificity (RMS) and prove that sentences from MSR satisfy RMS and the set of Maximum Specific Rules (MSR) is consistent.

2. Laws

Let $\mathfrak{L}$ be the first-order logic with signature $\Im = \langle P_1, ..., P_m \rangle$, m > 0, where $P_1, ..., P_m$ are the predicate symbols of arity $\mathrm{n_1 ..., n_m}$. An empirical system [5] is taken to mean a finite model $\mathfrak{M} = \langle \mathrm{B,W} \rangle$ of the signature $\Im$, where B is the basic set of the empirical system, and $\mathrm{W} = \langle \mathrm{P_1,...,P_m} \rangle$ is the tuple of predicates of the signature $\Im$ defined on B. Let $\mathrm{Th}(\mathfrak{M})$ be the set of all rules that are true on empirical system $\mathfrak{M}$ and has the form:

$$C = (A_1 \& ... \& A_{\mathrm{k}} \Rightarrow A_0),\ k\ \geq\ 0 \tag{2}$$

where $A_0, A_1, ..., A_k$ are literals. A literal is a predicate symbol or its negation with variables instantiated for arguments.

Proposition 2.1. *The rule* $C = (A_1 \& ... \& A_{\mathrm{k}} \Rightarrow A_0)$ *logically follows from any rule of the form:*

$$(A_{\mathrm{i1}} \& ... \& A_{\mathrm{ih}} \Rightarrow A_0), \{A_{\mathrm{i1}}, ..., A_{\mathrm{ih}}\} \subset \{A_1, ..., A_{\mathrm{k}}\},\ 0 \leq h < k, \tag{3}$$
$$\textit{that is}\ \ (A_{\mathrm{i1}} \& ... \& A_{\mathrm{ih}} \Rightarrow A_0) \vdash (A_1 \& ... \& A_{\mathrm{k}} \Rightarrow A_0).$$

Definition 2.1. By *subrule* of the rule $\mathrm{C} = (\mathrm{A_1 \& ... \& A_k \Rightarrow A_0})$ we mean any logically stronger rule of the form (3).

Corollary 2.1. *If a subrule of the rule C is true on* $\mathfrak{M}$*, then the rule C is also true on* $\mathfrak{M}$*.*

Definition 2.2. By the *law* on $\mathfrak{M}$, we mean any rule C of the form (2) that satisfies the following conditions [6]:

(1) C is true on $\mathfrak{M}$;

(2) the premise of the rule is not always false on $\mathfrak{M}$;
(3) none of its subrules is true on $\mathfrak{M}$.

Let $\mathcal{L}$ be the set of all laws on $\mathfrak{M}$. From the logic and methodology of science it is known that those hypotheses are laws that are most refutable, simple and contain the minimal number of the parameters. In our case, all these properties, that are usually difficult to define, follow from the deductive power of the laws. The 'subrules' are (i) logically stronger than the rules and more prone to become false (falsifiable) because they contain weaker premises and, therefore, applicable to bulkier data; (ii) simpler as containing less number of atomic expressions than the rule; (iii) including a smaller number of 'parameters' (the number of atomic expressions may be regarded as parameters 'tuning' the rules to data).

Theorem 2.1. *$\mathcal{L} \vdash Th(\mathfrak{M})$.*

3. The Probability of Events and Sentences

Let us generalize the notion of the law into the probabilistic case. For this purpose we introduce the probability on the model $\mathfrak{M}$. For the sake of simplicity we will follow paper [7], and introduce the probability μ as a discrete function on B, μ: B $\rightarrow$ [0,1], such that

$$\sum_{a \in B} \mu(a) = 1, \; and \; \mu(a) \neq 0, \; a \in B; \; \mu(D) = \sum_{b \in D} \mu(b), \; D \subseteq B \qquad (4)$$

We define the probability μ on the product of B^n as a probability function $\mu^n(a_1, ..., a_n) = \mu(a_1) \times ... \times \mu(a_n)$ More general definitions of the probability function μ are considered in [7].

Let us define the interpretation of the language $\mathfrak{L}$ on the empirical system $\mathfrak{M} = \langle$B,W$\rangle$ as mapping I: $\Im \rightarrow$ W, which associates with every signature symbol $P_j \in \Im$, j = 1,...,m, the predicate P_j from W of the same arity. Let X = {x_1, x_2, x_3, ... } be the set of all variables of the language $\mathfrak{L}$. By the validation ν is meant the function ν: X $\rightarrow$ B, mapping variables into the set of objects B.

Let us define the probability for the sentences of the language $\mathfrak{L}$. Let U($\Im$) be the set of all atomic formulas of the language $\mathfrak{L}$; $\Re(\Im)$ is the set of all the sentences of the language $\mathfrak{L}$, obtained by the closure of the set U($\Im$) with respect to standard Boolean constructs $\&, \vee, \neg$. By the $\hat{\varphi}$, $\varphi \in \Re(\Im)$ we define the formula, where the predicate symbols of $\Im$ are substituted by the predicates of W via interpretation I and by the $\nu\hat{\varphi}$ we define the formula, where variables of the formula $\hat{\varphi}$ are substituted by the objects of

A via the validation ν. In particular, $\nu\hat{P}_{\mathrm{j}}(\mathrm{x}^{\mathrm{j}}_1,...,\mathrm{x}^{\mathrm{j}}_{\mathrm{nj}})^{\varepsilon\mathrm{j}} = \mathrm{P}_{\mathrm{j}}(\mathrm{a}_1,...,\mathrm{a}_{\mathrm{j}})^{\varepsilon\mathrm{j}}$, $\nu(\mathrm{x}^{\mathrm{j}}_1)$ $= \mathrm{a}_1$,..., $\nu(\mathrm{x}^{\mathrm{j}}_{\mathrm{nj}}) = \mathrm{a}_{\mathrm{j}}$. Let us define the probability η of the sentences of $\Re(\Im)$. If $\mathrm{x}_1,\ldots,\mathrm{x}_{\mathrm{n}}$ are all variables of the sentence $\varphi \in \Re(\Im)$, then

$$\eta(\varphi) = \mu^{\mathrm{n}}(\{(a_1,\ldots,a_{\mathrm{n}})\,|\,\nu\hat{\varphi} \textit{ is true on } \mathfrak{M}, \nu(x_1)=a_1,\ldots,\nu(x_{\mathrm{n}})=a_{\mathrm{n}}\}) \quad (5)$$

4. The probabilistic Laws on $\mathfrak{M}$

Let us revise the concept of the law on $\mathfrak{M}$ in terms of probability. We do it in such a way that the concept of the law on $\mathfrak{M}$ would be a particular case of this definition. The law on $\mathfrak{M}$ is such a true rule, which subrules are false on $\mathfrak{M}$ or in other words the law is such a true rule, that cannot be made simpler or logically stronger without losing truth. This property of the law "not to be simplified" allows stating the law not only in terms of truth but also in terms of probability.

For any rule $\mathrm{C} = (\mathrm{A}_1\& \ldots\&\mathrm{A}_{\mathrm{k}} \Rightarrow \mathrm{A}_0)$ we will define the conditional probability of the rule $\eta(\mathrm{C}) = \eta(\mathrm{A}_0/\mathrm{A}_1\& \ldots\&\mathrm{A}_{\mathrm{k}}) = \eta(\mathrm{A}_0\&\mathrm{A}_1\& \ldots\&\mathrm{A}_{\mathrm{k}})/\eta(\mathrm{A}_1\& \ldots\&\mathrm{A}_{\mathrm{k}})$.

Theorem 4.1. *For any rule $C = (A_1\&...\&A_{\mathrm{k}} \Rightarrow A_0)$, the following two conditions are equivalent:*

(1) the rule C is the law on $\mathfrak{M}$ that satisfies the properties (1),(2), and (3) of the definition 2.2;
(2) (a) $\eta(C) = 1$;
(b) $\eta(A_1\&...\&A_{\mathrm{k}}) > 0$;
(c) the conditional probability $\eta(C)$ of the rule C is strictly more than conditional probabilities of each of its subrules.

Proof. *(1)(1) $\leftrightarrow$ (2)(a).* The rule C is true on $\mathfrak{M}$ iff due to the property (5) of the probability η of sentences $\eta(\mathrm{C}) = 1$.

(1)(2) $\leftrightarrow$ (2)(b). The premise of the rule C is not always false on $\mathfrak{M}$ iff there exist a validation ν such that $\nu(\widehat{A_1\&...\&A_k})$ is true on $\mathfrak{M}$. Due to the property (4) of the probability μ and the property (5) of the probability η it means that $\eta(\mathrm{A}_1\&...\&\mathrm{A}_{\mathrm{k}}) > 0$.

(1)(3) $\leftrightarrow$ (2)(c). The the conditional probability $\eta(\mathrm{C})$ of the rule C is equal to 1. We need to proof that conditional probability of each of its subrules is strictly less then 1. Let us consider one of its subrule $(A_{\mathrm{i}1}\&...\&A_{\mathrm{ih}} \Rightarrow A_0), \{A_{\mathrm{i}1},...,A_{\mathrm{ih}}\} \subset \{A_1,...,A_{\mathrm{k}}\},\ 0 \le h < k$. This subrule is not true on $\mathfrak{M}$ iff due to the property (5) of the probability η its probability is strictly less then 1. □

This theorem gives us the equivalent definition of the law on $\mathfrak{M}$.

Definition 4.1. By a *probabilistic law* on $\mathfrak{M}$ *with conditional probability 1* is meant the rule $\mathrm{C} = (\mathrm{A}_1\& \ ...\&\mathrm{A}_\mathrm{k} \Rightarrow \mathrm{A}_0)$ of the form (2) satisfying the following conditions:

(1) η (C) = 1, $\eta(\mathrm{A}_1\& \ ...\& \ \mathrm{A}_\mathrm{k}) > 0$;
(2) conditional probability of the rule η(C) is strictly greater than conditional probabilities of each of its subrules.

The next corollary follows from the theorem 4.1.

Corollary 4.1. *The rule is a probabilistic law on $\mathfrak{M}$ with conditional probability 1 iff it is a law on $\mathfrak{M}$.*

Let us consider items 1 and 2 of the theorem 4.1 from the standpoint of the 'not to be simplified' law:

- A law is such a true on $\mathfrak{M}$ rule, that cannot be simplified or to become logically stronger without a loss of the truth.
- Any logically stronger subrule of the rule has strictly less conditional probability (less than 1), so the rule cannot be simplified without loosing the value 1 of the conditional probability.

A more general definition of the law follows from these formulations:

Definition 4.2. The *law* is such a rule of the form (2) based on the truth values, conditional probability or other evaluations of the sentences, which cannot be made logically stronger without reducing their values.

Therefore, we can define the probabilistic law for the more general case by omitting the condition $\eta(\mathrm{C}) = 1$ from the point (1) of the definition 4.1.

Definition 4.3. By a *probabilistic law* on $\mathfrak{M}$, we designate such a rule $\mathrm{C} = (\mathrm{A}_1\& \ ...\&\mathrm{A}_\mathrm{k} \Rightarrow \mathrm{A}_0)$, of the form (2), the conditional probability of which is defined and strictly more than the conditional probabilities of each of its subrules. For a particular case of the subrule $\Rightarrow \mathrm{A}_0$ the conditional probability $\eta(\mathrm{C})$ of the rule C is strictly greater than the probability $\eta(\mathrm{A}_0)$.

Let us define by the $\mathcal{LP}$ the set of all probabilistic laws. It follows from the Theorem 4.1 and the definition 4.3 that the set $\mathcal{LP}$ includes the set $\mathcal{L}$.

Corollary 4.2. $\mathcal{L} \subset \mathcal{LP}$.

Definition 4.4. By the Strongest Probabilistic Law (SPL-rule) on $\mathfrak{M}$, we designate such a probabilistic law $\mathrm{C} = (\mathrm{A}_1\& \ ...\&\mathrm{A}_\mathrm{k} \Rightarrow \mathrm{A}_0)$, which is not a subrule of any other probabilistic law.

We define as SPL the set of all SPL-rules.

Proposition 4.1. *$\mathcal{L} \subset SPL \subset \mathcal{LP}$.*

5. Semantic Probabilistic Inference

Let us define the Semantic Probabilistic Inference of the set of laws $\mathcal{L}$ and the set of probabilistic laws $\mathcal{LP}$.

Definition 5.1. By the *Semantic Probabilistic Inference* (SP-inference) of the some SPL rule C we mean such a sequence of probabilistic laws, which we denote as the sequence $C_1 \sqsubset C_2 \sqsubset \cdots \sqsubset C_\mathrm{n}$, that:

$$\begin{gathered} C_1, C_2, \ldots, C_\mathrm{n} \in \mathcal{LP}, \\ C_\mathrm{i} = (A_1^i\& \ldots \&A_{ki}^i \Rightarrow G),\ i = 1, 2, \ldots n,\ n > 0, \\ the\ rules\ C_\mathrm{i}\ are\ subrules\ of\ the\ rules\ C_{\mathrm{i}+1}, \\ \eta(C_{\mathrm{i}+1}) > \eta(C_\mathrm{i}),\ i = 1, 2, \ldots, n-1, \\ C\ is\ C_\mathrm{n} \end{gathered} \tag{6}$$

Proposition 5.1. *Any probabilistic law from $\mathcal{LP}$ belongs to some SP-inference. For any SPL-rule there is some SP-inference of that rule.*

Corollary 5.1. *For any law from $\mathcal{L}$ there is some SP-inference of that law.*

Let us consider the set of all inferences of the sentence G. This set constitutes the Semantic Probabilistic Inference tree (SPI-tree) of this sentence.

Definition 5.2. By the maximum specific rule MS(G) for the I-S inference of the sentence G we mean the SPL rule of the SPI-tree of the sentence G, which has the maximum value of conditional probability.

We define as MSR the set of all maximum specific rules.

Proposition 5.2. *$\mathcal{L} \subset MSR \subset SPL \subset \mathcal{LP}$*

6. Probabilistic Maximum Specific Laws

Now we define the Requirement of Maximal Specificity (RMS). We will suppose that the class H of objects in (1) is defined by some sentence H

$\in \Re(\Im)$ of the language $\mathfrak{L}$. In this case the RMS says that p(G;H) = p(G;F) = r for this sentence. In terms of probability η it means that η(G/H) = η(G/F) = r for any H $\in \Re(\Im)$, satisfying (1).

Definition 6.1. The *Requirement of Maximal Specificity* (RMS):
if we add any sentence H $\in \Re(\Im)$ to the premise of the rule (F $\Rightarrow$ G), η(G/F) = r, such that F(**a**)&H(**a**) for some object **a**, then for the new rule (F&H $\Rightarrow$ G) we have η(G/F&H) = η(G/F) = r.

In other words the requirement RMS means that there is no other sentence H in $\Re(\Im)$ that increases (or decreases, see lemma 6.1 below) the conditional probability η(G/F) = r by adding it to the premise.

Lemma 6.1. *If the sentence H $\in \Re(\Im)$ decreases the probability η(G/F&H) < η(G/F) then the sentence ¬H increases it: η(G/F&¬H) > η(G/F).*

Proof. Let us denote a = η(G&F&H), b = η(F&H), c = η(G&F&¬H), d = η(F&¬H). Then the inequality η(G/F&H) < η(G/F) may be represented as a/b < (a+c)/(b+d). From the inequality a/b < (a+c)/(b+d) it follows that (a+c)/(b+d) < c/d $\Leftrightarrow$ η(G/F) < η(G/F&¬H) □

Lemma 6.2. *For any rule $C = (B_1\&...\&B_t \Rightarrow A_0)$, $\eta(B_1\&...\&B_t) > 0$, of the form (2) there is a probabilistic law $C' = (A_1\&...\&A_k \Rightarrow A_0)$ on $\mathfrak{M}$ which is subrule of the rule C and $\eta(C') \geq \eta(C)$.*

Theorem 6.1. *Any MS(G) rule satisfies the RMS requirement.*

Proof. We need to prove that for any sentence H $\in \Re(\Im)$ the equalities η(G/F&H) = η(G/F) = r take place for any MS(G) rule C = (F $\Rightarrow$ G). From the definition 6.1 it follows that there exists an object **a** such that F(**a**)&H(**a**). Due to the property (5) of the probability η we have that η(F&H) > 0 and, hence, the conditional probability is defined.

Let us consider the case when the sentence H is some atom B or its negation ¬B and η(G/F&H) $\neq$ r. Then, according to the lemma 6.1 one of the rules (F&B $\Rightarrow$ G), (F&¬B $\Rightarrow$ G) has the greater value of the conditional probability η(F&B $\Rightarrow$ G) > r or η(F&¬B $\Rightarrow$ G) > r. According to lemma 6.2 there exists a probabilistic law C', which is a subrule of the rule C and η(C') $\geq$ η(C) > r. The rule C' belongs to the SPI-tree and has the greater value of the conditional probability, that is contradict to the presupposition that C is MS(G) rule.

Let us consider the case when the sentence H is a conjunction of two atoms B_1&B_2 for which the theorem is true. If one of the inequalities η(G/F&B_1&B_2) > r, η(G/F&¬B_1&B_2) > r, η(G/F&B_1&¬B_2) > r, η(G/F&¬B_1&¬B_2) > r, takes place then according to lemma 6.2, there exists a probabilistic law C' ∈ SPI-tree, which is a subrule of the rule C and η(C') ≥ η(C) > r. This is impossible because C is a MS(G) rule. Hence, for all these inequalities we may have only equality = or inequality < . The last case is impossible due to the following equation

$$\frac{\eta(G\&F)}{\eta(F)} = r = \frac{GFB_1B_2}{FB_1B_2}, \quad where$$
$$GFB_1B_2 = \eta(G\&F\&B_1\&B_2) + \eta(G\&F\&\neg B_1\&B_2) +$$
$$\eta(G\&F\&B_1\&\neg B_2) + \eta(G\&F\&\neg B_1\&\neg B_2),$$
$$FB_1B_2 = \eta(F\&B_1\&B_2) + \eta(F\&\neg B_1\&B_2) +$$
$$\eta(F\&B_1\&\neg B_2) + \eta(F\&\neg B_1\&\neg B_2)$$

The case when the sentence H is a conjunction of some atoms or its negations may be proved by induction.

In general case the sentence H ∈ $\Re(\Im)$ may be presented as a disjunction of disjoint conjunctions of atoms and their negations. For completing the proof we need to consider the case when the sentence H is a disjunction of two disjoint sentences D∨E, η(D&E) = 0, for which the theorem is true and η(G/F&D) = η(G/F&E) = η(G/F) = r. It follows from the equation:

$$\eta(G/F\&(D\vee E)) = \frac{\eta(\mathrm{G\&F\&(D\vee E)})}{\eta(F\&(\mathrm{D\vee E}))} = \frac{\eta(\mathrm{G\&F\&D}) + \eta(\mathrm{G\&F\&E})}{\eta(F\&\mathrm{D}) + \eta(F\&\mathrm{E})} = r$$

The case of disjunction of more than two disjoint sentences is followed by induction from the case of two disjoint sentences. □

Corollary 6.1. *Any law on $\mathfrak{M}$ satisfies the RMS requirement.*

7. The solution of the statistical ambiguity problem

Theorem 7.1. *The I-S inference is consistent for any theory Th ⊂ MSR.*

Proof. Let us prove that for the sentences from Th ⊂ MSR it is impossible to obtain a contradiction when we have two inferences {A ⇒ G, B ⇒ ¬G} ⊂ Th ⊂ MSR, where η(A&B) > 0. We prove that in this case one of the

following rules is stronger (has a greater value of conditional probability) than the rules A $\Rightarrow$ G, B $\Rightarrow \neg$G.

$$A\&B \Rightarrow G,\ A\&B \Rightarrow \neg G,\ A\&\neg B \Rightarrow G,\ \neg A\&B \Rightarrow \neg G \tag{7}$$

Then, according to lemma 6.2, there exist probabilistic laws with conditional probability more then the rules A $\Rightarrow$ G, B $\Rightarrow \neg$G, which contradicts the condition Th $\subset$ MSR.

By contradiction the rules (7) have the conditional probability no more than the rules A $\Rightarrow$ G, B $\Rightarrow \neg$G.

(1) Let us consider the first rule A&B$\Rightarrow$G. By contradiction η(G/A&B) $\leq \eta$(G/A). Let us consider two cases:

(a) η (A&¬B) $\neq 0$. Since η(A&B) > 0, then

$$\eta(G/A) = \frac{\eta(A\&G)}{\eta(A)} =$$
$$\frac{\eta(A\&G\&B) + \eta(A\&G\&\neg B)}{\eta(A\&B) + \eta(A\&\neg B)} \geq \frac{\eta(A\&G\&B)}{\eta(A\&B)} \Leftrightarrow$$
$$\frac{\eta(A\&G\&\neg B)}{\eta(A\&\neg B)} \geq \mu(G/A) \geq \frac{\eta(A\&G\&B)}{\eta(A\&B)} \Leftrightarrow$$
$$\mu(G/A\&\neg B) \geq \mu(G/A) \geq \mu(G/A\&B)$$

If the first inequality is strong, then the other inequalities are also strong. Therefore from the inequality η(G/A&B) $<$ η(G/A) it follows that η(G/A&¬B) $> \eta$(G/A). It completes the proof for this case. The remaining case is η(G/A&B) $=$ η(G/A).

(b) η (A&¬B) $= 0$. Since η(A&B) > 0, then

$$\eta(G/A) = \frac{\eta(A\&G)}{\eta(A)} = \frac{\eta(A\&G\&B) + \eta(A\&G\&\neg B)}{\eta(A\&B) + \eta(A\&\neg B)} =$$
$$\frac{\eta(A\&G\&B)}{\eta(A\&B)} = \eta(G/A\&B)$$

The remaining case is the same η(G/A&B) $= \eta$(G/A).

(2) Let us consider the rule A&B $\Rightarrow \neg$G. By contradiction we have η(¬G/A& B) $\leq \eta$(¬G/B). By similar argumentation we have

$$\mu(\neg G/\neg A\&B) \geq \mu(\neg G/B) \geq \mu(\neg G/A\&B)$$

If the inequality η(¬G/A&B) $<$ η(¬G/B) is strong, then η(¬G/¬A&B) $> \eta$(¬G/B) and the theorem is proved for this case. The remaining case is η(¬G/A&B) $= \eta$(¬G/B).

(3) Let us consider the cases 1,2 when we have the equality:

μ (G / A& B) = μ (G / A)

μ($\neg$ G / A& B) = μ ($\neg$ G / B)

Then $\mu(G/A\&B) + \mu(\neg G/A\&B) = 1 = \mu(G/A) + \mu(\neg G/B)$

Since the rules A $\Rightarrow$ G and B $\Rightarrow$ $\neg$G are probabilistic laws and satisfy the conditions η($\neg$G/B) > η($\neg$G), η(G/A) > η(G). Then 1 = η(G/A) + η($\neg$G/B) > η(G) + η($\neg$G) = 1

We obtained the contradiction with the presupposition. □

Let us illustrate this theorem by the example of Jane Jones. We can define the maximum specific rules MS(E), MS($\neg$E) for the sentences E, $\neg$E as follows:

$\widehat{\text{L1}}$: 'Almost all cases of streptococcus infection, that are not resistant to streptococcus infection, clear up quickly after the administration of penicillin';

L2 : 'Almost no cases of penicillin resistant streptococcus infection clear up quickly after the administration of penicillin'.

The rule $\widehat{\text{L1}}$ has the greater value of conditional probability, than the rule L1 and, hence, it is a MS(E) rule for the sentences E. These two rules can't be fulfilled on the same data.

Conclusion: We can predict without contradictions if we use the set MSR as statistical laws in I-S inference.

8. The Relational Data Mining and program system 'Discovery'

Based on the semantic probabilistic inference the Relational Data Mining (RDM) approach to the intensive area of applications - Knowledge Discovery in Data Bases and Data Mining (KDD&DM) - was developed [8-10]. The program system 'Discovery', which utilizes this approach, has been implemented. In the frame of this approach we may discover the full (in the sense of theorem 2.1) and consistent (in the sense of theorem 7.1) set of rules. In [6] we argue that using RDM we may cognize the object domain. The system 'Discovery' realizes the Semantic Probabilistic Inference and can discover the sets of laws $\mathcal{L}$, $\mathcal{LP}$ and the sets SPL, MSR. The system 'Discovery' has been successfully applied to solving many practical tasks: cancer diagnostic systems, time series forecasting, psychophysics, bioinformatics, and many others (see www-site Scientific Discovery [11]).

Acknowledgments

The work is partially supported by the Russian Foundation for Basic Research 05-07-90185-v, Scientific Schools grant of the President of the Russian Federation 4413.2006.1.

References

1. Hempel, C. G. (1965) Aspects of Scientific Explanation, In: C. G. Hempel, *Aspects of Scientific Explanation and other Essays in the Philosophy of Science*, The Free Press, New York.
2. Hempel, C. G.: 1968, 'Maximal Specificity and Lawlikeness in Probabilistic Explanation', *Philosophy of Science* **35**, 116–33.
3. Salmon, W. C. (1990) *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis.
4. Yao-Hua Tan (1997) is default logic a reinvention of inductive-statistical reasoning? *Synthese* **110:** 357–379, *Kluwer Academic Publishers.*
5. Krantz, D.H., Luce, R.D., Suppes, P., Tversky, A. (1971, 1989, 1990), Foundations of measurement, Vol. 1,2,3 - NY, London: Acad. press, (1971) 577 p., (1989) 493 p., (1990) 356 p.
6. Evgenii Vityaev, Boris Kovalerchuk, Empirical Theories Discovery based on the Measurement Theory. *Mind and Machine*, v.14, # 4, 551-573, 2004
7. Halpern, J.Y. (1990), 'An analysis of first-order logic of probability', *Artificial Intelligence* **46**, pp.311-350.
8. Kovalerchuk, B., Vityaev, E. (2000), Data Mining in finance: Advances in Relational and Hybrid Methods, Kluwer Academic Publishers, 308 p.
9. Kovalerchuk, B., Vityaev, E., Ruiz, J.F. (2001), 'Consistent and Complete Data and "Expert" Mining in Medicine'. In: Medical Data Mining and Knowledge Discovery, Springer, pp.238-280.
10. Evgenii Vityaev, Boris Kovalerchuk. Data Mining For Financial Applications. In: O. Maimon and L. Rokach (eds.), Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, Springer 2005, pp.1203-1224.
11. Scientific Discovery http://www.math.nsc.ru/AP/ScientificDiscovery