

Studies in Computational Intelligence 796

María Eugenia Cornejo

László T. Kóczy

Jesús Medina

Antonio Eduardo De Barros Ruano

*Editors*

# Trends in Mathematics and Computational Intelligence



Springer

# How to Predict Consistently?



Evgeni Vityaev and Sergei Odintsov

**Abstract** One of reasons for arising the statistical ambiguity is using in the course of reasoning laws which have probabilistic, but not logical justification. Carl Hempel supposed that one can avoid the statistical ambiguity if we will use in the probabilistic reasoning maximal specific probabilistic laws. In the present work we deal with laws of the form  $\varphi \Rightarrow \psi$ , where  $\varphi$  and  $\psi$  are arbitrary propositional formulas. Given a probability on the set of formulas we define the notion of a maximal specific probabilistic law. Further, we define a prediction operator as an inference with the help of maximal specific laws and prove that applying the prediction operator to some consistent set of formulas we obtain a consistent set of consequences.

**Keywords** Probabilistic inference · Maximal specificity · Prediction Consistency

## 1 Introduction

The statistical ambiguity problem arises due to using in the course of reasoning laws which have a probabilistic, but not logical justification. Carl Hempel supposed that one can avoid the statistical ambiguity if we will use in the probabilistic reasoning only so called maximal specific probabilistic laws (see Sect. 2). In the present work we deal with laws of the form  $\varphi \Rightarrow \psi$ , where  $\varphi$  and  $\psi$  are arbitrary propositional formulas. In Sect. 3 we define the concept of probability on the set of formulas close to that of [1, 8] and extend it to the family of rules. Finally, in Sect. 4 we define the set of maximal specific probabilistic laws and the prediction operator as an inference with the help of maximal specific laws. Then we prove that applying the prediction

---

E. Vityaev · S. Odintsov (✉)  
Sobolev Institute of Mathematics, Novosibirsk, Russian Federation  
e-mail: odintsov@math.nsc.ru

E. Vityaev  
e-mail: vityaev@math.nsc.ru

© Springer Nature Switzerland AG 2019  
M. E. Cornejo et al. (eds.), *Trends in Mathematics and Computational Intelligence*, Studies in Computational Intelligence 796,  
[https://doi.org/10.1007/978-3-030-00485-9\\_4](https://doi.org/10.1007/978-3-030-00485-9_4)

operator to a consistent set of formulas we obtain the set of consequences, which is consistent too.

Despite the explanation and the prediction have the same logical structure, we prefer the term “prediction” in the following cases: if the act of prediction precedes a predicted fact in time. We also speak on a prediction if a predicted fact remains unknown due to different reasons: too high costs of establishing this fact, an impossibility to establish the fact at the moment, the lack of this fact in a series of preceding experiments, etc.

## 2 Statistical Ambiguity and Requirement of Maximal Specificity

The *Covering Law Model* suggested by Carl Hempel [3] (see [4, 7] for a historical overview) distinguished two kinds of explanation: Deductive-Nomological explanations (D-N explanations) and Inductive-Statistical explanations (I-S explanations). A D-N argument is a standard logical inference of facts from other facts with the help of general laws. An I-S argument has the form:

$$\frac{p(G; F) = t}{F(a)} \quad \frac{}{G(a)}$$

The line distinguishes the *explanandum*  $G(a)$  from two premisses (*explanans*), one of which has the form of a statistical law of the form  $p(G; F) = t$ ,  $0 \leq t \leq 1$ , where  $t$  denotes the probability that an object from the set defined by predicate  $F$  is also a member of the set defined by predicate  $G$ .

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 [4]. Recall one of traditional examples of the statistical ambiguity. Suppose that we have the following statements.

- L1 *Almost all cases of 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.*
- C1 *Jane Jones had streptococcus infection.*
- C2 *Jane Jones received treatment with penicillin.*
- C3 *Jane Jones had a penicillin resistant streptococcus infection.*

Following the above pattern one can construct two I-S explanations based on these statements. On the base of L1 and  $C1 \wedge C2$  one can explain why Jane Jones recovered quickly (E). The second argument with premisses L2 and  $C2 \wedge C3$  explains why Jane Jones did not ( $\neg E$ ). The set of premisses  $\{C1, C2, C3\}$  is consistent. However, the 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—they should contain all relevant information with respect to the domain in question. In our example, then, the premiss  $C3$  invalidates the first argument, since this argument is not maximally specific with respect to all information about Jane Jones. So, we can only explain  $\neg E$ , but not  $E$ .

In [4] Hempel defined the Requirement of Maximal Specificity (RMS) as follows. An I-S argument

$$\frac{p(G; F) = t}{F(a)} \quad \frac{F(a)}{G(a)}$$

is an acceptable I-S explanation with respect to a “knowledge state”  $K$ , if the following Requirement of Maximal Specificity is satisfied. For any predicate  $H$  for which the following two sentences are contained in  $K$ :  $\forall x(H(x) \Rightarrow F(x))$ ,  $H(a)$ , there exists a statistical law  $p(G; H) = t'$  in  $K$  such that  $t = t'$ . The basic idea of RMS is that if  $F$  and  $H$  both contain the object  $a$ , and  $H$  is a subset of  $F$ , then  $H$  provides more specific information about the object  $a$  than  $F$ , and therefore the law  $p(G; H)$  should be preferred over the law  $p(G; F)$ . However the law  $p(G; H)$  has the same probability as the law  $p(G; F)$ .

### 3 Probability on Propositional Formulas and Rules

Our attention will be restricted to propositional logic. We start from a set of atoms  $At$  and construct from them the set of well formed formulas  $F(At)$  using connectives  $\wedge, \vee, \rightarrow, \neg$ . As usual the equivalence  $\leftrightarrow$  is considered as an abbreviation. We define  $\top$  as  $\varphi \vee \neg\varphi$ , where  $\varphi$  is some fixed formula. For a finite set of formulas  $T$  the conjunction of its elements is denoted by  $\bigwedge T$ . By  $V(\varphi)$  we denote the set of atoms occurring in formula  $\varphi$ . The set  $F(At)$  with naturally interpreted connectives forms an algebra of formulas  $\mathcal{F}(At)$ . The classical interpretation of propositional connectives is assumed, therefore models for our logic can be identified with mappings from  $At$  to the set of classical truth values  $\{0, 1\}$ . We call such mappings ( $At$ )-valuations. Every valuation  $v : At \rightarrow \{0, 1\}$  extends in a standard way to the set  $F(At)$  using classical truth tables for connectives  $\wedge, \vee, \rightarrow$ , and  $\neg$ , the extended valuation we denote in the same way  $v : F(At) \rightarrow \{0, 1\}$ . A formula  $\varphi$  is *satisfiable* in a set  $\mathfrak{G}$  of valuations if  $v(\varphi) = 1$  for some  $v \in \mathfrak{G}$ , a formula  $\varphi$  *holds on*  $\mathfrak{G}$ ,  $\mathfrak{G} \models \varphi$ , if  $v(\varphi) = 1$  for all  $v \in \mathfrak{G}$ . Finally, a set  $T \subseteq F(At)$  is  $\mathfrak{G}$ -consistent if there is  $v \in \mathfrak{G}$  such that  $v(\varphi) = 1$  for all  $\varphi \in T$ . The set of all valuations we denote  $\mathfrak{All}$ .

Let  $\mathfrak{G} \subseteq \mathfrak{All}$ . The relation  $\varphi \equiv_{\mathfrak{G}} \psi$  is defined by the condition that  $\varphi \leftrightarrow \psi$  holds on  $\mathfrak{G}$ . It is clear that  $\equiv_{\mathfrak{G}}$  is a congruence on  $\mathcal{F}(At)$ , the respective quotient is denoted as  $\mathcal{B}^{\mathfrak{G}}(At)$ . The coset of  $\varphi$  w.r.t.  $\equiv_{\mathfrak{G}}$  is denoted as  $[\varphi]_{\mathfrak{G}}$ . Recall that the universe of  $\mathcal{B}^{\mathfrak{G}}(At)$  equals  $\{[\varphi]_{\mathfrak{G}} \mid \varphi \in F(At)\}$  and that this set is finite whenever  $At$  is finite. The operations of  $\mathcal{B}^{\mathfrak{G}}(At)$  are denoted as  $\wedge_{\mathfrak{G}}, \vee_{\mathfrak{G}}, \rightarrow_{\mathfrak{G}}, \neg_{\mathfrak{G}}$  and the lattice order as  $\sqsubseteq_{\mathfrak{G}}$ . Recall that  $[\varphi]_{\mathfrak{G}} \sqsubseteq_{\mathfrak{G}} [\psi]_{\mathfrak{G}}$  iff  $[\varphi]_{\mathfrak{G}} = [\varphi \wedge \psi]_{\mathfrak{G}} = [\varphi]_{\mathfrak{G}} \wedge_{\mathfrak{G}} [\psi]_{\mathfrak{G}}$ .

Let  $\mu : 2^{\mathfrak{G}} \rightarrow [0, 1]$  be a finitely additive measure defined on  $\mathfrak{G}$ , i.e.,  $\mu$  is such that: (1)  $\mu(\mathfrak{G}) = 1$ ; 2)  $\mu(\emptyset) = 0$ ; and (3)  $\mu(A_1 \cup \dots \cup A_n) = \mu(A_1) + \dots + \mu(A_n)$  for pairwise disjoint subsets  $A_1, \dots, A_n \subseteq \mathfrak{G}$ . Additionally we assume that  $\mu(A) = 0$  implies  $A = \emptyset$ . Elements of  $\mathfrak{G}$  may be interpreted as outcomes of experiments. Further, we assume that only essential experiments are included in  $\mathfrak{G}$ , which explains why  $\mu(\{v\}) \neq 0$  for all  $v \in \mathfrak{G}$ .

For every  $\varphi \in F(At)$ , put  $\varphi^{\mathfrak{G}} := \{v \in \mathfrak{G} \mid v(\varphi) = 1\}$ . Now we define a function  $\nu : F(At) \rightarrow [0, 1]$  by the rule  $\nu(\varphi) = \mu(\varphi^{\mathfrak{G}})$ . It is easy to check that  $\nu$  satisfies the following properties.

1.  $\nu(\varphi) = 1$  iff  $\varphi$  is holds on  $\mathfrak{G}$ .
2.  $\nu(\varphi) = 0$  iff  $\varphi$  is not satisfiable on  $\mathfrak{G}$ .
3.  $\nu(\varphi \vee \psi) = \nu(\varphi) + \nu(\psi)$  iff  $\varphi \wedge \psi$  is not satisfiable in  $\mathfrak{G}$ .

This means that  $\nu$  is a *probability* on the set of propositional formulas in a sense close to that of [1, 8].

Now we generalize the notions from [9]. By a *rule* we mean a syntactic object of the form

$$r = \varphi \Rightarrow \psi,$$

where  $\varphi, \psi \in F(At)$  and  $\varphi \rightarrow \psi$  is not a logical tautology. We call  $\varphi$  and  $\psi$  a *body* and a *head* of  $r$ :  $\varphi = B(r)$  and  $\psi = H(r)$ . The probability of a rule  $r = \varphi \Rightarrow \psi$  with  $\nu(\varphi) \neq 0$  is defined as follows:  $\nu(r) := \nu(\psi | \varphi) = \frac{\nu(\psi \wedge \varphi)}{\nu(\varphi)}$ . In case,  $\varphi$  is not satisfiable on  $\mathfrak{G}$ , the value  $\nu(r)$  remains undefined. Notice that the value  $\nu(r)$  was defined so that it is smaller or equal to  $\nu(\varphi \rightarrow \psi)$ .

**Definition 1** Let  $r_1$  and  $r_2$  be two rules with the same head,  $H(r_1) = H(r_2)$ . We call  $r_2$  a *generalization* of  $r_1$ , symbolically  $r_2 \succeq r_1$  if  $B(r_1)^{\mathfrak{G}} \subseteq B(r_2)^{\mathfrak{G}}$ ; rule  $r_2$  is a *proper generalization* of  $r_1$ ,  $r_2 \succ r_1$ , if  $r_2 \succeq r_1$  and  $B(r_1)^{\mathfrak{G}} \neq B(r_2)^{\mathfrak{G}}$ . We say in this case that  $r_1$  is a (*proper*) *specialization* of  $r_2$ .

In other words, one of the two rules with the same head is a proper generalization of the other if its body is weaker from the logical point of view.

## 4 Prediction Operator

In this section we generalize the results of [10].<sup>1</sup> Let  $At$ , a set  $\mathfrak{G}$  of valuations, and a measure  $\mu$  on  $G$  be fixed. Assume that  $\mathcal{R}$  is a set of rules such that for  $r \in \mathcal{R}$  and a rule  $s$  such that  $s \prec r$  and  $\nu(s) > \nu(r)$ , there is  $r' \in \mathcal{R}$  with  $r' \preceq s$ . Now we introduce two special subsets of  $\mathcal{R}$ :

$$\mathbf{M}_1(\mathcal{R}) = \{r \in \mathcal{R} \mid (\top \Rightarrow H(r)) \succ r \text{ implies } \nu(r) > \nu(\top \Rightarrow H(r))\};$$

<sup>1</sup>In [10], rules are of the form  $\alpha_1 \wedge \dots \wedge \alpha_n \Rightarrow \beta$ , where  $\alpha_1, \dots, \alpha_n, \beta$  are literals, i.e., atoms or negations of atoms.

$$\mathbf{M}_2(\mathcal{R}) = \{r \in \mathcal{R} \mid r \in \mathbf{M}_1(\mathcal{R}) \text{ and } \forall r' \in \mathbf{M}_1(\mathcal{R})(r \succ r' \Rightarrow v(r') \leq v(r))\}.$$

Namely rules from  $\mathbf{M}_2(\mathcal{R})$  we consider as satisfying the Requirement of Maximal Specificity, because a specification of such rule by a rule from  $\mathbf{M}_1(\mathcal{R})$  does not lead to an increase of probability, which means that for  $r \in \mathbf{M}_2(\mathcal{R})$ , its body contains all statistically relevant information for the prediction of  $H(r)$ .

For a set of rules  $\Pi \subseteq \mathbf{M}_2(\mathcal{R})$  we define an operator of *direct predictions*:

$$Pr_\Pi(T) = T \cup \{H(r) \mid r \in \Pi, \exists \varphi_1, \dots, \varphi_n \in T (\mathfrak{G} \models (\varphi_1 \wedge \dots \wedge \varphi_n) \leftrightarrow B(r))\},$$

where  $T$  is a set of formulas. Further, we put:

$$Pr_\Pi^0(T) = T, \quad Pr_\Pi^{n+1}(T) = Pr_\Pi(Pr_\Pi^n(T)), \quad PR_\Pi(T) = \bigcup_{n \in \omega} Pr_\Pi^n(T).$$

It is clear that  $PR_\Pi(T)$  is the least fixed point of the operator of direct predictions containing  $T$ . We call  $PR_\Pi$  a *prediction operator* for  $\Pi$ .

**Theorem 1** *Let  $At$  be a finite set of atoms,  $\Pi \subseteq \mathbf{M}_2(\mathcal{R})$ , and  $T \subseteq F(At)$ . If  $T$  is a  $\mathfrak{G}$ -consistent set of formulas, then  $PR_\Pi(T)$  is  $\mathfrak{G}$ -consistent.*

*Proof* Obviously, it will be enough to check that the operator of direct predictions produces a  $\mathfrak{G}$ -consistent set of formulas. Further, since  $At$  is finite, the family of all formulas is finite up to equivalence, and we may assume that sets  $T$  and  $\Pi$  are finite too.

Let  $\Pi' = \{r_0, \dots, r_n\}$  be the set of such rules  $r$  from  $\Pi$  that the equivalence  $(\varphi_1 \wedge \dots \wedge \varphi_n) \leftrightarrow B(r)$  holds on  $\mathfrak{G}$  for some  $\varphi_1, \dots, \varphi_n \in T$ . We put  $T_0 = T$ ,  $T_{i+1} = T_i \cup \{H(r_i)\}$ . Clearly,  $T_n = PR_\Pi(T)$ . Now we prove by induction that every  $T_i$  is  $\mathfrak{G}$ -consistent.

Assume that  $T_i$  is  $\mathfrak{G}$ -consistent, but  $T_{i+1}$  is not. Let  $r_i = \varphi \Rightarrow \psi$ . By definition of  $\Pi'$  there are  $\chi_1, \dots, \chi_n \in T$  such that  $(\chi_1 \wedge \dots \wedge \chi_n) \leftrightarrow \varphi$  holds on  $\mathfrak{G}$ . Let  $N = T_i \setminus \{\chi_1, \dots, \chi_n\}$ . Assume that  $\{\varphi, \neg(\bigwedge N)\}$  is  $\mathfrak{G}$ -consistent, i.e.,  $v(\varphi \wedge \neg(\bigwedge N)) \neq 0$ . In this case for  $s = \varphi \wedge \neg(\bigwedge N) \Rightarrow \psi$  we have:

$$v(s) = \frac{v(\varphi \wedge \neg(\bigwedge N) \wedge \psi)}{v(\varphi \wedge \neg(\bigwedge N))} = \frac{v(\varphi \wedge \psi) - v(\varphi \wedge \bigwedge N \wedge \psi)}{v(\varphi) - v(\varphi \wedge \bigwedge N)}.$$

We have  $\mathfrak{G} \models \bigwedge T_{i+1} \leftrightarrow (\varphi \wedge \bigwedge N \wedge \psi)$  and  $\mathfrak{G} \models \bigwedge T_i \leftrightarrow (\varphi \wedge \bigwedge N)$  by choice of  $\chi_1, \dots, \chi_n$ . Since by assumption  $v(\bigwedge T_{i+1}) = 0$  and  $v(\bigwedge T_i) \neq 0$ , we conclude that  $v(\varphi \wedge \bigwedge N \wedge \psi) = 0$  and  $v(\varphi \wedge \bigwedge N) \neq 0$ . In this way, we have

$$v(s) = \frac{v(\varphi \wedge \psi)}{v(\varphi) - v(\varphi \wedge \bigwedge N)} > \frac{v(\varphi \wedge \psi)}{v(\varphi)} = v(r_i).$$

Since  $v(s) > v(r_i)$  there is  $r' \in \mathcal{R}$  such that  $r' \preceq s$ . It follows from  $v(r') > v(r_i)$  and  $r_i \in M_1(\mathcal{R})$  that  $r' \in M_1(\mathcal{R})$ . On the other hand, from  $r_i \in M_2(\mathcal{R})$  and  $r_i > r'$  we obtain  $v(r_i) \geq v(r')$ . This contradiction proves that the body of  $s$  is not  $\mathfrak{G}$ -consistent:  $v(\varphi \wedge \neg(\bigwedge N)) = 0$ . As a consequence we obtain  $v(\varphi \wedge \neg(\bigwedge N) \wedge \psi) = 0$ . Now we have:

$$v(\varphi \wedge \psi) = v(\varphi \wedge \psi) - v(\varphi \wedge \neg(\bigwedge N) \wedge \psi) = v(\varphi \wedge \bigwedge N \wedge \psi) = 0.$$

Thus,  $v(r_i) = 0$ . At the same time  $r_i \in M_1(\mathcal{R})$ , which implies  $0 = v(r_i) > v(\top \Rightarrow \psi) \geq 0$ . The obtained contradiction concludes the proof.

Putting  $\mathcal{R}$  to be the family of all rules trivializes the above statement, because in this case the prediction operator turns into a consequence operator over  $\mathfrak{G}$ . Refinement Theorem [10] shows that the class of rules considered in [10] satisfies the requirements imposed on  $\mathcal{R}$ . Other non-trivial cases of  $\mathcal{R}$  will be considered in subsequent papers.

Of course, we worked in the ideal situation assuming that the probability on the set of formulas is known. In reality we have only statistical approximation of probabilities. The concept of semantical probabilistic inference (see [9]) aided at the search for maximal specific rules (of the form  $\bigwedge_{i=1}^n \alpha_i \Rightarrow \beta$ , where  $\alpha_i$  and  $\beta$  are literals) on the base of statistically verified data gives a well-working approximation of  $M_2(\mathcal{R})$ -rules. This search procedure was realized in the program system Discovery [5]. The description of applications of this system to financial forecasting and to medicine can be found in [5, 6].

**Acknowledgements** The first of the authors (Sects. 1 and 2, also a coauthor of Theorem 1) was supported by the Russian Science Foundation (project # 17-11-01176). Both authors are grateful to the anonymous referees for their helpful reports and to participants of ESCIM'17 for the interesting discussion.

## References

1. Fagin, R., Halpern, J.Y., Megiddo, N.: A logic for reasoning about probabilities. Inform. Comput. **80**, 78–128 (1990)
2. Fetzer, J.H.: Scientific Explanation. D. Reidel, Dordrecht (1981)
3. Fetzer, J.: Carl Hempel. In: Zalta, E.N. (ed.) Stanford Encyclopedia of Philosophy. Stanford University (2014). <https://plato.stanford.edu/entries/hempel>
4. Hempel, C.G.: Aspects of scientific explanation. In: Hempel, C.G. (ed.) Aspects of Scientific Explanation and other Essays in the Philosophy of Science. The Free Press, New York (1965)
5. Kovalerchuk, B., Vityaev, E.: Data Mining in Finance: Advances in Relational and Hybrid methods, 308 pp. Kluwer Academic Publishers (2000)
6. Kovalerchuk, B., Vityaev, E., Ruiz, J.F.: Consistent and complete data and “expert” mining in medicine. In: Medical Data Mining and Knowledge Discovery, pp. 238–280. Springer (2001)
7. Salmon, W.C.: Four Decades of Scientific Explanation. University of Minnesota Press, Minneapolis (1990)

8. Scott, D., Krauss P.: Assigning probabilities to logical formulas. In: Hintikka, J., Suppes, P. (eds.) *Aspects of Inductive Logic*, pp. 219–264. North-Holland (1966)
9. Vityaev, E.E.: The logic of prediction. In: *Proceedings of the 9th Asian Logic Conference*, Novosibirsk, Russia, 16–19 August 2006, pp. 263–276. World Scientific (2006)
10. Vityaev, E.E., Martynovich, V.V.: Probabilistic formal concepts with negation. In: Voronkov, A., Virbitskaite, I. (eds.) *PSI 2014. LNCS*, vol. 8974, pp. 1–15. Springer (2015)