KYBERNETIKA — VOLUME 24 (1988), NUMBER 6

A NOTE ON NONAXIOMATIZABILITY OF INDEPENDENCE RELATIONS GENERATED BY CERTAIN PROBABILISTIC STRUCTURES

IVAN KRAMOSIL

There exist such systems of random variables that the set of all true assertions, concerning the statistical independence of subsystems of random variables and expressed in the terms of a relational calculus, cannot be derived from an effectively decidable (i.e. recursive) set of axioms by the usual deduction rules.

1. INTRODUCTION

Consider a probabilistic structure $Q = \{X_i\}_{i \in I}$, i.e. a system of random variables parametrized by a parametric set I, with each X_i taking an abstract probability space $\langle \Omega, \mathscr{S}, \mu \rangle$ into a measurable space $\langle Y_i, \mathscr{A}_i \rangle$. From the probability theory point of view, the complete characterization of the structure Q is given by its *simultaneous probability (distribution)* P_I, defined on $\langle \mathbf{X}_{i \in I} Y_i, \mathbf{X}_{i \in I} \mathscr{A}_i \rangle$ when setting, for each system $\{E_i\}_{i \in I}$, $E_i \in \mathscr{A}_i$, of sets

$$\mathsf{P}_I(\mathbf{X}_{i \in I} E_i) = \mu(\bigcap_{i \in I} \{\omega : \omega \in \Omega, X_i(\omega) \in E_i\}) . \tag{1}$$

For every $\emptyset \neq A \subset I$ the *marginal probability* P_A is defined in the same way, just replacing $\{X_i\}_{i \in I}$ by its subsystem $\{X_i\}_{i \in A}$. Evidently, for each $\emptyset \neq A \subset I$ and each system $\{E_i\}_{i \in A}$,

$$\mathsf{P}_A(\mathbf{X}_{i \in A} E_i) = \mathsf{P}_I(\mathbf{X}_{i \in I} F_i) , \quad F_i = E_i \quad \text{for} \quad i \in A , \qquad F_i = Y_i \quad \text{for} \quad i \in I - A . \tag{2}$$

Pairwise disjoint subsets $\{A_i\}_{i \in B}$, $\emptyset \neq A_i \subset I$, are called *statistically (stochastically) independent*, if the substructures $Q_{A_i} = \{X_j\}_{j \in A_i}$ of Q are statistically (stochastically) independent, i.e., for each $E_i \in \mathbf{X}_{j \in A_i} \mathscr{A}_j$, $\mathsf{P}_{\bigcup_{i \in B} A_i}(\mathbf{X}_{i \in B} E_i) = \prod_{i \in B} \mathsf{P}_{A_i}(E_i)$. The structure Q_A is *completely independent*, if all singletons of A are statistically independent; in such a case the system of marginal probabilities defined by all singletons of A defines uniquely the simultaneous probability P_A.

If card $Y_i = m$ for each $i \in I$ and card $I = n$, then the computation of P_I requests, in the worst case, to compute $\mathsf{P}_I(x)$ for m^n values x from $\mathbf{X}_{i \in I} Y_i$, and this problem is known to be NP-complete. On the other side, in the case of complete independence just $m \,.\, n$ values must be obtained (the values $\mathsf{P}_{\{i\}}(y)$ for each $i \in I$ and $y \in Y_i$). Supposing that the values for P_I are not computed, but estimated from random samples, the situation in the general case is still much worse and we usually have not data enough to estimate $\mathsf{P}_I(x)$ for all $x \in \mathbf{X}_{i \in I} Y_i$ at a reasonable confidence level. Hence, the potential statistical independence of some subsets of I may significantly reduce the computational complexity connected with the computation or estimation of P_I and may also simplify decision making for the purpose of which P_I is to be obtained.

The problem to decide, whether two sets $A, B \subset I$, arr statistically independent or not is theoretically very simple, at least for finite sets I and Y_i. The only we have to do is to verify the equality $\mathsf{P}_{A \cup B} = \mathsf{P}_A \mathsf{P}_B$ for all possible argument values, but this may be rather complicated, either from the computational complexity point of view, or in case the values of $\mathsf{P}_{A \cup B}$, P_A and P_B are to be statistically estimated. So the idea has been suggested (cf. [3], e.g.) to find such logical relations among the statistical dependences of various pairs $A, B \subset I$, that having verified the independence for some pairs A, B, the independence of other pairs could be derived by purely logical means, i.e. without references to the values of the corresponding marginal probability measures. The aim of this paper is to show that in spite of some non-negligible successes achieved in this direction, this goal cannot be satisfied in general, i.e., for certain probabilistic structures Q the corresponding relation of statistical independence cannot be completely described in such a "deductive" way.

2. VALIDITY AND PROBABILITY OF STATISTICAL INDEPENDENCE RELATIONS

First of all, we need a formal language $\mathscr{L}_{\mathscr{I}}$ to describe the statistical independence relation generated by a probabilistic structure $Q = \{X_i\}_{i \in I}$.

Individual indeterminates (we use this term in order to keep the expression "variable" free for the context "random variable") are $A, B, C, \ldots$, possibly with indices, and they are interpreted as to range over the space $\mathscr{P}(I)$ of all subsets of I.

Individual constants are $\bar{I}$, the name of the parameter set I, $\emptyset$, the name of the empty subset of I, and, supposing that I is finite or countable, then for each $i \in I$, $\{i\}$ is also an individual constant – the name of the singleton subset of I containing just the element i.

Two binary functions are $\cup$ and $-$ with their usual set-theoretic interpretation (set union and set difference), more set-theoretic functions can be defined a posteriori.

Terms are defined in the usual inductive way: individual indeterminates and individual constants are terms and, supposing that s and t are terms, $s \cup t$ and

$s - t$ are also terms (the same being valid for the other functions possibly defined through the two basic ones).

The only *relation of the language* $\mathscr{L}_{\mathscr{I}}$ is the binary relation $\mathscr{I}$ hence, *elementary formulas* are of the form $\mathscr{I}(s, t)$, where s and t are terms. Elementary formulas are formulas, if F and G are formulas and A is an indeterminate, then $\neg F$ (negation of F), $F \vee G$ (disjunction) and $(\forall A) F$(for all A. . .) are formulas, other connectives and existential quantifier may be extra-defined in the common way. Let $L(\mathscr{I})$ denote the set of all formulas of the language $\mathscr{L}_{\mathscr{I}}$.

Each probabilistic structure $Q = \{X_i\}_{i \in I}$ generates a *model* or *interpretation* $\mathscr{M}(Q)$ of the language $\mathscr{L}_{\mathscr{I}}$ as follows.

The support $M(Q)$ of $\mathscr{M}(Q)$, i.e. the set in which the indeterminates take their values, is the set $\mathscr{P}(I)$ of all subsets of the parameter space I.

Individual constants $\tilde{I}$, $\emptyset$ and $\{i\}$, $i \in I$ are interpreted as the particular subsets of I as mentioned above, also binary functions $\cup$ and $-$ are interpreted as the above introduced set operations. Hence, each term of $\mathscr{L}_{\mathscr{I}}$ can be unambiguously interpreted as a subset of I supposing that all indeterminates occurring in the term are interpreted as subsets of I.

Let s, t be terms of $\mathscr{L}_{\mathscr{I}}$, let φ be a mapping (evaluation) ascribing a subset of I to each indeterminate occurring in s or in t, let $\varphi(s)$, $\varphi(t)$ be the subsets of I ascribed to s and t by this evaluation. Then the elementary formula $\mathscr{I}(s, t)$ is defined *to be satisfied* in $\mathscr{M}(Q)$ (by $\mathscr{M}(Q)$, to *hold* in $\mathscr{M}(Q)$, to be *valid* or *true* in $\mathscr{M}(Q)$) *with respect to* φ, if $\varphi(s)$ and $\varphi(t)$ are statistically independent subsets of I $(=I(Q))$, in symbols

$$(3) \qquad \mathscr{M}(Q), \varphi \models \mathscr{I}(s, t) \Leftrightarrow \mathsf{P}_{\varphi(s) \cup \varphi(t)} = \mathsf{P}_{\varphi(s)} \mathsf{P}_{\varphi(t)} .$$

The definition of meta-relation $\models$ is extended to other formulas of $\mathscr{L}_{\mathscr{I}}$ according to the Tarski classical definition of truth predicate (cf. [4]). I.e., for E, $F \in L(\mathscr{I})$, for an indeterminate A and for an evaluation φ of indeterminates from E and F, $\mathscr{M}(Q), \varphi \models \neg E$ iff $\mathscr{M}(Q), \varphi \models E$ does not hold, $\mathscr{M}(Q), \varphi \models E \vee F$ iff either $\mathscr{M}(Q)$, $\varphi \models E$ or $\mathscr{M}(Q), \varphi \models F$, finally, $\mathscr{M}(Q), \varphi \models (\forall A) E$ iff $\mathscr{M}(Q), \tilde{\varphi} \models E$ for all evaluations $\tilde{\varphi}$ ascribing the same values as φ to all indeterminates from E other than A. The rules for other connectives and for the existential quantifier can be easily completed. Hence, if $E \in L(\mathscr{I}) \cap Sent$, i.e. if E is a sentence (a formula without free indeterminates) of $\mathscr{L}_{\mathscr{I}}$, then either $\mathscr{M}(Q) \models E$ or $\mathscr{M}(Q) \models \neg E$. Let us define the set of true sentences of $\mathscr{L}_{\mathscr{I}}$ with respect to Q by

$$(4) \qquad Tr(\mathscr{L}_{\mathscr{I}}, Q) = \{E : E \in L(\mathscr{I}) \cap Sent, \mathscr{M}(Q) \models E\} .$$

This expression defines a subset of $L(\mathscr{I})$ in a *semantical* way, referring to the meanings ascribed to the symbols of $\mathscr{L}_{\mathscr{I}}$ and to the relation of statistical independence defined in $\mathscr{P}(I) \times \mathscr{P}(I)$ through the probabilistic structure Q. *Syntactically* defined subsets of $L(\mathscr{I})$ are those defined through the relation of logical consequence. If Ax is a subset of $L(\mathscr{I})$, then $Cn(Ax) \subset L(\mathscr{I})$ denotes the subset of all formulas which are derivable from formulas in Ax by the usual deduction rules (say, modus ponens, generalization and substitution into propositional tautologies). A subset $Ax \subset L(\mathscr{I})$ of formulas is called *recursive*, if its characteristic function defined on $L(\mathscr{I})$ is recursive, i.e., Ax is recursive, if there is an algorithm (effective procedure) deciding for each $E \in L(\mathscr{I})$, whether $E \in Ax$ or not.

Now, the relation $\mathscr{I} = \mathscr{I}(Q)$ of statistical independence, generated by a probabilistic structure $Q = \{X_i\}_{i \in I}$, is defined to be *axiomatizable*, if there exists a recursive subset $Ax \subset L(\mathscr{I})$ of axioms such that

$$Sent \cap Cn(Ax) = Tr(\mathscr{L}_{\mathscr{I}}, Q)\,. \tag{5}$$

The following assertion is almost trivial.

Fact 1. If Q is a probabilistic structure with a finite parameter space I, then the corresponding statistical independence relation $\mathscr{I}$ is axiomatizable.

Proof. If I is finite, then every subset $A \subset I$ can be expressed by the term $\bigcup_{i \in A} \{i\}$ of $\mathscr{L}_{\mathscr{I}}$. There are $2^{2 \cdot \mathrm{card}\, I}$ elementary formulas $\mathscr{I}(s, t)$ with different pairs of term values $\langle s, t \rangle$ and syntactical equality of terms is decidable. Moreover, each formula of the form $(\forall A)\, E$ is equivalent to the finite conjunction $\bigwedge_{S \in \mathscr{P}(I)} E(\bar{S})$, where $E(\bar{S})$ results from E by replacing A with the term $\bar{S} = \bigcup_{i \in S} \{i\}$, and this equivalence is decidable. Hence, every formula from $Tr(\mathscr{L}_{\mathscr{I}}, Q)$ is decidably equivalent to a formula from a recursive set of formulas from $L(\mathscr{I})$ without quantifiers, so that $Tr(\mathscr{L}_{\mathscr{I}}, Q)$ is also recursive and can serve, trivially, as a set of axioms. □

Fact 1 cannot be extended, in general, to infinite parameter spaces. This negative assertion follows almost immediately from the existence of nonaxiomatizable binary relations over infinite domains. However, for the reader's convenience, and in order to make him familiar with a standard proof technique, rather seldom applied in the field of probability theory and mathematical statistics, we shall devote the next chapter to a more detailed construction and proof.

3. THE MAIN ASSERTION

First of all, let us introduce the notion of Peano arithmetic (PA) which will be of crucial importance in what follows. The language $\mathscr{L}_{PA}$ of PA contains an infinite sequence $x, y, z, x_1, y_1, z_1, \ldots$ of individual indeterminates, one individual constant 0 (zero), two binary function symbols $+$ and $.$ (addition and multiplication), one unary function symbol S (successor), and the binary relation $=$ of equality. The elementary terms $+\,(x, y)$ and $.\,(x, y)$ will be written in the usual infix form $x + y$ and $x \,.\, y$, other terms are defined inductively in the usual way. Elementary formulas are $s = t$ (again, the infix form is preferred) for terms s and t, other formulas are also defined inductively in the usual way. Axioms of PA are those of pure first-order predicate calculus, those for equality relation (reflexivity, symmetry, transitivity), and the following ones:

(A1) $$(\forall x)(\forall y)(S(x) = S(y) \Rightarrow x = y),$$

(A2) $$\neg(\exists x)(S(x) = 0),$$

(A3) $$(\forall x)(x + 0 = x),$$

(A4) $$(\forall x)(\forall y)(x + S(y) = S(x + y)),$$

(A5) $$(\forall x)(x \, . \, 0 = 0),$$

(A6) $$(\forall x)(\forall y)(x \, . \, S(y) = (x \, . \, y) + x),$$

finally, for each formula $\psi(x)$ of $\mathscr{L}_{PA}$ we have the following induction axiom

(A7) $$(\psi(0) \wedge (\forall x)(\psi(x) \Rightarrow \psi(S(x)))) \Rightarrow (\forall x)\, \psi(x).$$

The deduction rules are the usual ones of the first-order predicate calculus.

An interpretation of PA is defined by setting a set N ranged by the indeterminates (the elements of N are called natural numbers), by defining two binary operations, addition and multiplication, on N, by defining a unary operation of successor on N and by setting the zero element 0 in N. The equality relation is interpreted as identity in N. Every interpretation of PA together with an evaluation of indeterminates defines uniquely the set of valid elementary formulas of the form $s = t$ with terms s and t. By the usual inductive way described above we define the set of all sentences of PA which are true under the interpretation in question. Consider only such interpretations under which all axioms of PA introduced above are true formulas and denote by $Tr(PA)$ the set of all formulas of $\mathscr{L}_{PA}$ which are true under all interpretations with this property. The following fact is nothing else than a reformulation of the famous Gödel incompleteness theorem.

Fact 2. There is no recursive set Ax of formulas of $\mathscr{L}_{PA}$ such that $Cn(Ax) = Tr(PA)$.

Having borrowed the greatest Gödel's result, we shall not hesitate to make still another profit of his results introducing the notion of Gödel numbers or Gödel enumerations in their classical simple form. Each formula of $\mathscr{L}_{PA}$ is a finite sequence of symbols (letters) from a finite alphabet, supposing that indexed indeterminates are taken as subsequences consisting of x, y, z and of numerals 0, 1, ..., 9. Considering an ordering of these elementary symbols, we may ascribe to each symbol a its (positive) index $q(a)$ with respect to this ordering. Let $p_1, p_2, \ldots$ be the sequence of all prime numbers in the increasing order, i.e. $p_1 = 2$, $p_2 = 3$, $p_3 = 5, \ldots$ To each finite sequence $\alpha = \alpha_1 \alpha_2 \ldots \alpha_n$ of symbols we may ascribe its Gödel number $gn(\alpha)$, setting

(6) $$gn(\alpha) = gn(\alpha_1 \alpha_2 \ldots \alpha_n) = \prod_{j=1}^{n} p_j^{q(\alpha_j)}.$$

Evidently, given a formula of $\mathscr{L}_{PA}$, its Gödel number can be effectively computed. On the other hand, given a positive integer n, we may effectively decompose it into the product of primes in corresponding powers, we may inspect, whether all primes up to some p_n occur in this expression and if it is the case, we may verify, whether the sequence of corresponding powers yields a formula of $\mathscr{L}_{PA}$ or not. Hence, we may effectively verify, whether a given positive integer n is the Gödel number of a formula of $\mathscr{L}_{PA}$ or not. Finally, set

$$G = GN(Tr(PA)) = \{gn(\alpha): \alpha \in Tr(PA)\}\ , \tag{7}$$

i.e. G is the set of Gödel numbers of all true sentences of PA. Now, we have at hand everything we need to state and prove the main assertion.

Theorem. There exists a probabilistic structure Q with an infinite countable parameter space I such that the corresponding statistical independence relation $\mathscr{I}$ is not axiomatizable.

Proof. Take $I = \{0, 1, 2, \ldots\}$ and consider a sequence $Q_G = \{X_i\}_{i=0}^{\infty}$ of random variables defined on an abstract probability space $\langle \Omega, \mathscr{S}, \mu \rangle$ taking their values in the binary set $\{0, 1\}$ and satisfying the following conditions:

$$\mu(\{\omega: \omega \in \Omega, X_0(\omega) = 0\}) = \mu(\{\omega: \omega \in \Omega, X_0(\omega) = 1\}) = \tfrac{1}{2}\ , \tag{8}$$

if $j \in G$, then the random variable X_j is statistically independent of X_0 and equally distributed as X_0,

if $j \in I - G$, then the corresponding conditional probabilities read

$$\mu(\{\omega: \omega \in \Omega, X_j(\omega) = 0\}/\{\omega:\omega \in \Omega, X_0(\omega) = 0\}) = p\ , \tag{9}$$

$$\mu(\{\omega: \omega \in \Omega, X_j(\omega) = 0\}/\{\omega: \omega \in \Omega, X_0(\omega) = 1\}) = 1 - p \tag{10}$$

for some p, $\frac{1}{2} < p < 1$, with complementary probabilities for $X_j(\omega) = 1$ in both the cases. Hence, random variables X_0 and X_j are statistically independent iff $j \in G$, in other way written $\mathscr{M}(Q_G) \models \mathscr{I}(\{0\}, \{j\})$ iff $j \in G$, hence, $\mathscr{M}(Q_G) \models \mathscr{I}(\{0\}, \{gn(\varphi)\})$ iff $\varphi \in Tr(PA)$.

However, we have arrived at this conclusions by a deductive way, hence, our constructions concerning the probabilistic system Q_G and the Gödel enumeration gn can be converted into a finite set of axioms Ax_0 such that ($Ax_0 \vdash \psi$ means that $\psi \in Cn(Ax_0)$)

$$Ax_0 \vdash \mathscr{I}(\{0\}, \{gn(\varphi)\}) \Leftrightarrow gn(\varphi) \in G\ , \tag{11}$$

and

$$Ax_0 \vdash (gn(\varphi) \in G) \Rightarrow \varphi\ . \tag{12}$$

Now, supposing that the relation $\mathscr{I}$ generated by Q_G were axiomatizable, we have a recursive set of axioms $Ax_1 \subset \mathscr{L}_{\mathscr{I}}$ such that

$$Ax_1 \vdash \mathscr{I}(\{0\}, \{gn(\varphi)\}) \tag{13}$$

iff $\varphi \in Tr(PA)$. Hence, taking the set $Ax_0 \cup Ax_1$ of axioms (and rewriting the axioms from Ax_1 into the language $\mathscr{L}_{PA}$), we obtain a recursive set Ax of axioms such that $Ax \vdash \varphi$ if $\varphi \in Tr(PA)$, hence, $Cn(Ax) = Tr(PA)$ and we have arrived at a contradiction with Fact 2, so that no axiomatization Ax_1 for $\mathscr{I} = \mathscr{I}(Q_G)$ exists. □

4. COMMENTS AND CONCLUSIONS

The proof just presented deserves, perhaps, some more comments. The existence of a set G such that the predicate $x \in G$ is axiomatizable in the framework of a given formalized language can be proved more simply and more directly from the cardinality reasonings. The set of well-formed formulas of each formalized language, as a set of finite sequences over finite or countable alphabet, is also countable. A recursive set of axioms is uniquely defined by a recursive function taking the set of all formulas into $\{0, 1\}$ and ascribing the value 1 just to the axioms. The number of recursive functions is countable (each recursive function is defined by a finite program over a finite alphabet), hence, there are only countably many axiomatic systems within the given language and, consequently, the set of subsets of $I = \{0, 1, 2, \ldots\}$, for which the membership predicate is axiomatizable, is also countable. However, the space $\mathscr{P}(I)$ of all subsets of I is not countable, so that the existence of $G \subset I$ with nonaxiomatizable membership predicate immediately follows (in order to take a particular G with this property for further considerations, the application of the axiom of choice is inevitable). In a sense, almost all sets from $\mathscr{P}(I)$ have the same property as G. Or, let V be a random variable taking a probability space $\langle \Omega, \mathscr{S}, \mu \rangle$ into $\mathscr{P}(I)$, such that $\mu(\{\omega: \omega \in \Omega, V(\omega) = E\}) = 0$ for all $E \in \mathscr{P}(I)$, then evidently (an appropriate σ-field is supposed to be defined on $\mathscr{P}(I)$)

$$\mu(\{\omega: \omega \in \Omega, \ulcorner x \in V(\omega) \urcorner \textit{ is not axiomatizable}\}) = 1 . \tag{14}$$

The random variable V can be easily defined, say, by a sequence $V_0, V_1, V_2, \ldots$ of independent and identically distributed random variables, taking $\langle \Omega, \mathscr{S}, \mu \rangle$ into $\{0, 1\}$ and such that $0 < \mu(\{\omega: \omega \in \Omega, V_i(\omega) = 1\}) < 1$, the only we have to do is to set

$$V(\omega) = \{j: j \in I = \{0, 1, 2, \ldots\}, V_j(\omega) = 1\} . \tag{15}$$

On the contrary to this "existential" consideration, in the proof presented above a particular case of the set G is given which is, in spite of the nonaxiomatizability of its membership relation, definable in semantical way.

Let us close this paper by a short reconsidering of Fact 1. The existence of an axiomatization of the independence relation for finite parameter spaces follows almost trivially, on the other hand, the problem to find a shorter more sophisticated axiomatization (than those given by an exhaustive listing of non-equivalent quantifier-less true formulas) is far from being trivial and may be of theoretical and practical significance. The same holds when discussing about the aim to find some common properties possessed by all independence relations generated by probabilistic structures in question, or at least by probabilistic structures with finite parameter spaces; some interesting results in this direction have been achieved, e.g., by Studený [5] and Matúš [2]. Such general properties of statistical independence relations may serve for a meta-axiomatization of the notion of statistical independence, but cannot provide a complete axiomatization for all particular independence relations generated by particular probabilistic structures. Hence, there is no general algorithm yielding axiomatizations for all probabilistic structures parametrized by finite sets of integers just as particular instances of an axiomatization for an infinite parameter space, So, for infinitely many cases of probabilistic structures with finite parameter space, a creative effort will be necessary to obtain an axiomatization of the corresponding independence relation.

The paper is almost self-explanatory up to references dealing with the Gödel's incompleteness theorem (Fact 2), the details can be found in any textbook on mathematical logic and we introduce [4] as the probably most accessible one in our country. Also the most elementary references on probability theory can be found in any textbook, let us mention [1] or [6].

(Received January 18, 1988.)

REFERENCES

[1] W. Feller: An Introduction to Probability Theory and its Applications, Vol. I. Second edition. J. Wiley & Sons — Chapman & Hall, New York—London 1957 (Russian translation: Mir, Moscow 1964).

[2] F. Matúš: Axiomatizations of the notion of independence — private communication.

[3] J. Pearl: On the logic of probabilistic dependencies. TR 4/86, Cognitive Systems Laboratory, Dept. of Comp. Sci., Univ. of California, Los Angeles.

[4] J. R. Shoenfield: Mathematical Logic. Addision-Wesley Publ. Comp. 1967 (Russian translation: Nauka, Moscow 1975).

[5] M. Studený: Multiinformation and the problem of characterization of conditional independence relations. Problems Control Inform. Theory *19* (1989), No. 1 (to appear).

[6] J. Štěpán: Teorie pravděpodobnosti (Probability Theory). Academia, Praha 1987.

RNDr. Ivan Kramosil, DrSc., Ústav teorie informace a automatizace ČSAV (Institute of Information Theory and Automation — Czechoslovak Academy of Sciences), Pod vodárenskou věží 4, 182 08 Praha 8. Czechoslovakia.