A NEW MODEL OF COMBINATORIAL PROBABILITY

RADIM JIROUŠEK

A new model of probability generalizing combinatorial approach is introduced. It complies with some practical demands on modelling comparative probability for finite sets useful for application in ES. Though only initial ideas and a simple characterization is presented the approach seems to be promising with respect to practical feasibility.

1. INTRODUCTION

An application of the probability theory to uncertainty management in expert systems has many advantages. One can profit by making use of both theoretical results and practical algorithms which were achieved by a number of outstanding mathematicians in the past. But still, there are situations when application of the classical (i.e. Kolmogorov axiomatic) probability model is rather improper and when it would be better to use some other model. For example, it involves situations when we have no statistical estimates of any probability and its subjective estimates are too vague.

In this paper we shall generalize a rather unfamiliar Kolmogorov's combinatorial model (introduced in [1]) in such a way that it acquires comparative capability. Moreover, the described approach is convenient for direct application in computer programs being very promising with respect to algorithmical complexity of its implementation.

2. *C*-DISTRIBUTIONS

Let us consider a finite set S of elementary events. In a standard combinatorial model a probability of any event $A \subset S$ is defined by the following ratio

$$\mathsf{P}(A) = \frac{\text{Number of element. events belonging to } A}{\text{Number of all elementary events}} = \frac{\operatorname{card}(A)}{\operatorname{card}(S)}$$

and a conditional probability of an event $A \subset S$ given an event $B \subset S$ equals

$$\mathsf{P}(A \mid B) = \frac{\operatorname{card}(A \cap B)}{\operatorname{card}(B)}\,.$$

Therefore, there is no possibility to compare positive conditional probabilities of elementary events s_1, s_2 given an event $B \subset S$ as it always holds

$$\mathsf{P}(\{s_1\} \mid B) > 0 \;\&\; \mathsf{P}\{s_2\} \mid B) > 0 \Rightarrow \mathsf{P}(\{s_1\} \mid B) = \mathsf{P}(\{s_2\} \mid B) = \frac{1}{\operatorname{card}(B)}\,.$$

To improve declarative capabilities of the model we assign to each elementary event $s \in S$ a nonnegative integer $p(s)$ in such a way that $p(s) = 0$ if and only if s is an impossible event and $p(s_1) > p(s_2)$ declares that the event s_1 is more probable than the event s_2. But, an equality $p(s_1) = p(s_2)$ should not be interpreted as equality of probabilities of the events s_1 and s_2 but rather as an impossibility to decide (or an ignorance) which of them is higher.

Definition 1. Let S be a finite set of elementary events (card $(S) \geqq 2$) and K be an integer, $K \geqq 2$. A C_K-distribution on S is any function

$$p: S \rightarrow \{0, 1, 2, \ldots, K - 1\} .$$

For any $K \geqq 2$ the class of all C_K-distributions on S will be denoted $C_K(S)$. It is evident that

$$C_K(S) \subset C_{K+1}(S) .$$

As we mostly will not be interested in exact values of K for which $p \in C_K(S)$ we will speak simply about C-distributions denoting

$$C(S) = \bigcup_{K=2}^{\infty} C_K(S) .$$

To exclude trivialities we shall suppose for all C-distributions p $\sum_{s \in S} p(s) > 0$ in the sequel.

3. SIMILARITY OF C-DISTRIBUTIONS

Definition 2. C-distributions $p, \bar{p} \in C(S)$ are equivalent ($p \equiv \bar{p}$ in notation) if there exists a positive number r such that

$$p(s) = r \,.\, \bar{p}(s)$$

for all $s \in S$. C-distributions $p, \bar{p}$ are similar ($p \sim \bar{p}$) if

(i) $p(s) = 0 \Leftrightarrow \bar{p}(s) = 0$ and

(ii) $[p(s_1) < p(s_2) \Rightarrow \bar{p}(s_1) \leqq \bar{p}(s_2)]$ & $[\bar{p}(s_1) < \bar{p}(s_2) \Rightarrow p(s_1) \leqq p(s_2)]$
for any pair $s_1, s_2 \in S$.

The above definition introduces a nontransitive relationship. Consider e.g. $S = \{s_1, s_2, s_3\}$ and three C-distributions p_1, p_2, p_3 defined on S:

$$p_1(s_1) = 1 , \quad p_1(s_2) = 2 , \quad p_1(s_3) = 3 ,$$
$$p_2(s_1) = 1 , \quad p_2(s_2) = 1 , \quad p_2(s_3) = 3 ,$$
$$p_3(s_1) = 2 , \quad p_3(s_2) = 1 , \quad p_3(s_3) = 3 .$$

It is evident that $p_1 \sim p_2$ and $p_2 \sim p_3$, but, p_1 and p_3 are not similar.

We shall not discuss this property in detail in this place, however, it should be mentioned that the property becomes advantageous when inconsistent pieces of knowledge are integrated in an expert system. Such a process of knowledge integration can be based on the idea of finding a C-distribution similar to all input pieces of knowledge (supposing they are expressed in a form of C-distributions). But, naturally, there may exist several such C-distributions. For example, the distribution p_4 defined on $S = \{s_1, s_2, s_3\}$

$$p_4(s_1) = p_4(s_2) = p_4(s_3) = 1$$

is (as well as p_2) similar to both p_1 and p_3 from the above example. Therefore, we need a tool enabling to express a measure of similarity. For this purpose we can use a function

$$D(p; q) = \sum_{s\in S} p(s) \log \left(\sum_{s\in S} q(s) \right) - \sum_{s\in S} p(s) \log q(s)$$

derived from the well-known Shannon relative entropy. Using the known properties of this entropy we can deduce the following simple assertions. Formulating them S is a finite set (with at least two elements) and K, L are integers greater than 1.

Theorem 1. Let $p, q, \bar{q} \in C(S)$. If $q \equiv \bar{q}$ then $D(p; q) = D(p; \bar{q})$.

Proof. The equality follows from the definition of the equivalence of C-distributions. Denote r the positive number for which $q(s) = r \cdot \bar{q}(s)$ holds for all $s \in S$. Then one can compute

$$\begin{aligned} D(p; q) &= \sum p(s) \log \left(\sum q(s)\right) - \sum p(s) \log q(s) = \\ &= \sum p(s) \log \left(\sum r\, \bar{q}(s)\right) - \sum p(s) \log r\, \bar{q}(s) = \\ &= \sum p(s) \log \left(\sum \bar{q}(s)\right) + \sum p(s) \log r - \sum p(s) \log \bar{q}(s) - \sum p(s) \log r = \\ &= D(p; \bar{q}) . \end{aligned}$$

□

Theorem 2. Let $p \in C_K(S)$ and $\bar{p} \in C_L(S)$. If $p \equiv \bar{p}$ then

$$D(p; \bar{p}) = \min_{q\in C_L(S)} (D(p; q)) .$$

Proof. The assertion follows immediately from the properties of the Shannon relative entropy. Let us define classical distributions P and Q on S by

$$P(\bar{s}) = p(\bar{s}) / \sum_{s\in S} p(s) , \quad Q(\bar{s}) = q(\bar{s}) / \sum_{s\in S} q(s) .$$

It is known that the Shannon relative entropy

$$H(P; Q) = \sum_{s\in S} P(s) \log \frac{P(s)}{Q(s)}$$

is nonnegative and equals 0 iff $P = Q$.

Elementary computations show that

$$D(p; q) = \sum_{s\in S} p(s)\, H(P; Q) + \sum_{s\in S} p(s) \log \left(\sum_{s\in S} p(s) \right) - \sum_{s\in S} p(s) \log p(s) ,$$

and therefore, if $p \equiv q$ then $P = Q$, $H(P; Q) = 0$ and $D(p; q)$ is minimal. □

Theorem 3. Let $p \in C_K(S)$ and $\bar{p} \in C_L(S)$. If

$$D(p; \bar{p}) = \min_{q\in C_L(S)} (D(p; q))$$

then $p \sim \bar{p}$. Moreover, if $L \geqq K$ then $p \equiv \bar{p}$.

Proof. If $p(s) > 0$ and $\bar{p}(s) = 0$ for an $s \in S$ then $D(p; \bar{p}) = +\infty$ which contradicts the assumption. Therefore $\bar{p}(s) = 0 \Rightarrow p(s) = 0$ for all $s \in S$. If we define a C_L distribution q

$$q(s) = \bar{p}(s) \quad \text{if} \quad p(s) > 0\,,$$
$$q(s) = 0 \quad \text{if} \quad p(s) = 0\,,$$

we get

$$D(p; q) = D(p; \bar{p}) - \sum_{s \in S} p(s) \log \frac{\sum_{s \in S} \bar{p}(s)}{\sum_{s \in S} q(s)}\,.$$

Since it is supposed that $D(p; \bar{p})$ is minimal

$$\sum_{s \in S} q(s) = \sum_{s \in S} \bar{p}(s)\,.$$

must hold. Thus we have proved that

$$p(s) = 0 \Leftrightarrow \bar{p}(s) = 0\,.$$

Now, let us suppose that there exists a couple $s_1, s_2 \in S$ such that

$$p(s_1) < p(s_2) \;\&\; \bar{p}(s_1) > \bar{p}(s_2)\,.$$

Analogously to the previous part of the proof we define

$$q(s) = \bar{p}(s) \quad \text{if} \quad s_1 \neq s \neq s_2\,,$$
$$q(s_1) = \bar{p}(s_2)\,,$$
$$q(s_2) = \bar{p}(s_1)\,.$$

Since

$$\sum_{s \in S} q(s) = \sum_{s \in S} \bar{p}(s)$$

we get

$$\begin{aligned} D(p; q) = {} & D(p; \bar{p}) + p(s_1) \log \bar{p}(s_1) + p(s_2) \log \bar{p}(s_2) - \\ & - p(s_1) \log q(s_1) - p(s_2) \log q(s_2) = \\ = {} & D(p; \bar{p}) + [p(s_1) - p(s_2)] \,.\, [\log (\bar{p}(s_1)/\bar{p}(s_2))] < D(p; \bar{p}) \end{aligned}$$

which contradicts the assumption that $D(p; \bar{p})$ is minimal.

Up to now, we have proved the similarity of p and $\bar{p}$. The last statement of the theorem follows from the inclusion $C_K(S) \subset C_L(S)$ and from the fact that $H(P; Q)$ minimizes only when $P = Q$. □

4. EXAMPLE

In this section we shall illustrate on a very simple example how the introduced apparatus can be used for inference in decision support systems.

Let us consider 3 predicates A, B, C describing possible results when casting of a die (with numbers 1 – 6)

A ...... the result is greater than 3
B ...... the result is odd
C ...... the result is 5.

With help of them we can define several finite sets of elementary events. Define S_0 consisting of 8 elementary events

$$S = \{ABC, AB\bar{C}, A\bar{B}C, \ldots, \bar{A}\bar{B}C, \bar{A}\bar{B}\bar{C}\}$$

and S_1, S_2, S_3 consisting of only 4 events each

$$S_1 = \{AC, A\bar{C}, \bar{A}C, \bar{A}\bar{C}\}\,,$$
$$S_2 = \{BC, B\bar{C}, \bar{B}C, \bar{B}\bar{C}\}\,,$$
$$S_3 = \{AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}\}\,.$$

For every C-distribution q defined on S_0 we can define a triplet q_1, q_2, q_3 of C-distributions defined on S_1, S_2, S_3 respectively according to the following scheme:

$$q_1(AC) = q(ABC) + q(A\bar{B}C)\,,$$
$$q_1(A\bar{C}) = q(AB\bar{C}) + q(A\bar{B}\bar{C})\,,$$
$$\ldots\ldots\ldots\ldots$$
$$q_2(BC) = q(ABC) + q(\bar{A}BC)\,,$$
$$\ldots\ldots\ldots$$
$$q_3(\bar{A}\bar{B}) = q(\bar{A}\bar{B}C) + q(\bar{A}\bar{B}\bar{C})\,.$$

Let us suppose that the input pieces of knowledge describe all dependencies between any two predicates. They are expressed in a form of three C-distributions defined on S_1, S_2, S_3

	A	$\bar{A}$
C	1	0
$\bar{C}$	2	3

P_1

	B	$\bar{B}$
C	1	0
$\bar{C}$	2	3

P_2

	B	$\bar{B}$
A	1	2
$\bar{A}$	2	1

P_3

with an interpretation that

$$\text{Prob.}\,(\bar{A}C) = \text{Prob.}\,(\bar{B}C) = 0\,,$$
$$\text{Prob.}\,(\bar{A}\bar{C}) > \text{Prob.}\,(A\bar{C}) > \text{Prob.}\,(AC)\,,$$
$$\text{Prob.}\,(\bar{B}\bar{C}) > \text{Prob.}\,(B\bar{C}) > \text{Prob.}\,(BC)\,,$$
$$\text{Prob.}\,(\bar{A}\bar{B}) < \text{Prob.}\,(A\bar{B}) > \text{Prob.}\,(AB)\,,$$
$$\text{Prob.}\,(\bar{A}\bar{B}) < \text{Prob.}\,(\bar{A}B) > \text{Prob.}\,(AB)\,.$$

In order to be able to determine any predicate when values of the two others is known we are looking for a C-distribution q defined on S_0 which induces C-distributions q_1, q_2, q_3 (on S_1, S_2, S_3 respectively) similar with C-distributions p_1, p_2, p_3 respectively.

It is evident that for any such q

$$q(A\bar{B}C) = q(\bar{A}BC) = q(\bar{A}\bar{B}C) = 0\,.$$

Denoting all others values of q according to the table

	B		$\bar{B}$	
	A	$\bar{A}$	A	$\bar{A}$
C	a	0	0	0
$\bar{C}$	b	c	d	e

we get from the definition of the similarity the following inequalities

$$a \leqq b + d \leqq c + e$$
$$a \leqq b + c \leqq d + e$$
$$a + b \leqq d$$
$$a + b \leqq c$$
$$e \leqq d$$
$$e \leqq c$$

There is an infinite number of C-distributions on S_0 meeting these conditions. Taking $\sum_{i=1}^{3} D(p_i; q_i)$ as a criterion the following distribution is the optimal one

	B		$\bar{B}$	
	A	$\bar{A}$	A	$\bar{A}$
C	1	0	0	0
$\bar{C}$	0	2	2	1

which really describes the underlying example best. Ftom this C-distribution q one can deduce e.g.

$$q(AB\bar{C}) = 0\,, \quad q(ABC) = 1$$

corresponding to the fact

$$\text{Prob.}\,(C \mid AB) = 1$$

or

$$q(A\bar{B}\bar{C}) = q_3(A\bar{B}) = 2\,, \quad q(\bar{A}\bar{B}\bar{C}) = q_3(\bar{A}\bar{B}) = 1$$

which can be interpreted

$$\text{Prob.}\,(A \mid \bar{B}\bar{C}) > \text{Prob.}\,(\bar{A} \mid \bar{B}\bar{C})$$

and

$$\text{Prob.}\,(A \mid \bar{B}) > \text{Prob.}\,(\bar{A} \mid \bar{B})\,.$$

REFERENCE

[1] A. N. Kolmogorov: Three approaches to a definition of the concept "quantity of information". Problems Inform. Transmission *1* (1965), 3—6.

Radim Jiroušek, CSc., 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.