PREPRINTS

of the

STOCHASTIC CONTROL SYMPOSIUM

BUDAPEST, HUNGARY

September 25—27, 1974

FUZZY SYSTEM APPROACH TO DECISION-MAKING BASED ON PUBLIC OPINION INVESTIGATION THROUGH QUESTIONNAIRES

P.Barnev, V.Dimitrov and P.Stanchev

Institute of Mathematics and Mechanics of Bulgarian Academy of Sciences
Sofia - Bulgaria

ABSTRACT

Questionnaire systems represent the stochastic feed-back, which is absolutely necessary when an effective control of human beings coallitions is to be realized. As usual, a man gives his answers to the questions put to him by an inquery, unprecisely and fuzzy. The more complicated the questions, the more general, unclear and fuzzy the human answers. An efficient questionnaire system for public opinion investigation must be able to deal with the fuzzy human answers in such a way that maximum information necessary for making a certain decision should be obtained from them. In the proposed paper a method based on the fuzzy system theory of Zadeh, is carried out. The method gives an optimal transformation of the fuzzy data obtained as a result of the public opinion investigation into structure, quite suitable for decision-making. A simple algorithm for computer realization of the considered method and some practical examples illustrated how the method works are given and discussed as well.

INTRODUCTION

The systems for public opinion investigation or questionnaire systems represent a stochastic feed-back that is absolutely necessary for effective governing of groups of people. The inaccurate and potentially uncertain and indeterminate environment where the every day actions of people take place has flooded human vocabulary with such words as "likely", "probably", "possibly", "perhaps", "maybe", "approximately", "more or less", "almost", "nearly", "as though", "so to say", "to all appearances", etc. Usually a man uses these words when formulating his answers to questions put to him. The more complex the questions are, the more unclear and fuzzier become man's answers. To deal with such kind of answers is beyond the capacities of a system based on categorical "black and white" reasonings of the type "yes-not" or on logical formalisms supposing that in the end every formula leads to the two alternatives "true" and "false" only. An effective questionnaire system must know how to manipulate with freely formulated fuzzy answers by the inquired men, in order to obtain from them as much information as possible, necessary for taking a decision. The advantages received as a result of the doing away with restrictions on the answers of the inquired men are evident. In the former questionnaire systems the inquired man's functions have been reduced to putting a cross or a circle round either one or another answer planned beforehand by the inquirer (questionnaire organizer). In the system manipulated with freely formulated answers

- the sketchness and triteness of the generated answers are avoided;
- the rise of descredited questionnaire realization (as a result of incorrect planning of answers) is minimized;
- the individual specific features of information given by the questionnaire participators are preserved to the highest degree.

On figure 1 a simplified sketch of management of group of people placed in a real inaccurate and potentially uncertain environment is shown. The questionnaire system is presented as a necessary intermediate link that ensures an optimal functioning of block of control (governing block).

FUZZY SUBSETS, LINGUISTIC EXPRESSIONS AND THEIR QUANTITATIVE ESTIMATES

Before describing the system for processing of the answers received by questionnaire realization, the problem of finding quantitative estimates of given verbal combination (linguistic expressions) is considered. For solving this problem the fuzzy system theory approach of Zadeh [1] is used.

A subset U_* of given set of elements U can be determined by means of binary function $b(u)$ defined for any $u \in U$ as

$$b(u) = \begin{cases} 1 \text{ if } u \in U_* \\ 0 \text{ if } u \notin U_* \end{cases}$$

In a number of cases the strict definition of a specified subset of given set of elements is hampering; for example - a determination of the subset of "adult men" of the set of "all people". In these cases it is convenient to use the concept of fuzzy set defined as follows.

Let U be a set of objects (in general, nonnumerical) and $f(u)$ be a numerical function such that $0 \le f(u) \le 1$ for any $u \in U$.

A set U_f of all ordered pairs $(f(u), u)$, $u \in U$ is called conventionally a fuzzy subset of U, corresponded to the function f and is represented as

$$U_f = f \mid U$$

f is called a membership function. In a case if U is a finite or countable with elements $u_1, u_2, \ldots$ then U_f is denoted by

$$U_f = (f_1 \mid u_1, f_2 \mid u_2, \ldots)$$

where $f_i = f(u_i)$

Every membership function f denotes some fuzzy subset of U. The value f takes for a given element $u \in U$ is a numerical characteristic (measure) of the grade of membership ("closeness") of U to the fuzzy subset U_f.

The operation NOT and POWER α ($\alpha > 0$) on a given fuzzy subset U_f and AND or OR on two given subsets U_f and U_g ($U_f, U_g \subset U$) are defined as follows:

$$\text{NOT}: \quad \bar{U}_f = (1-f) \mid U \qquad (1)$$

$$\text{POWER}: \quad U_f^{\alpha} = f^{\alpha} \mid U, \ \alpha > 0 \qquad (2)$$

$$\text{AND}: \quad U_f \wedge U_g = \min(f,g) \mid U \qquad (3)$$

$$\text{OR}: \quad U_f \vee U_g = \max(f,g) \mid U \qquad (4)$$

If the set U is a finite or countable one, i.e. $U = \{u_1, u_2, \ldots\}$, then:

$$U_f = \{(1-f_1) \mid u_1, (1-f_2) \mid u_2, \ldots\} \qquad (1)$$

$$U_f^{\alpha} = \{f_1^{\alpha} \mid u_1, f_2^{\alpha} \mid u_2, \ldots\} \qquad (2)$$

$$U_f \wedge U_g = \{\min(f_1, g_1) \mid u_1, \ldots\} \qquad (3)$$

$$U_f \vee U_g = \{\max(f_1, g_1) \mid u_1, \ldots\} \qquad (4)$$

Every linguistic expression represents a primary term or a combination of primary terms, the negation "not", the connectives "and" and "or" and some linguistic hedges. (Only combinations admissible in the natural languages are considered).

The primary terms are labels of specified fuzzy subsets (e.g. "young", "old", "large", "small", "clever", "strong", "pretty", "capable", etc.).

The linguistic hedges express some quantitative characteristics of the primary terms (e.g. "very", "much", "little", "slightly", "more", "less", "more or less", "scarcely", "substantially", "unconsiderably", "enough", etc.).

The fuzzy subset (i.e.) its membership function) corresponded to a given linguistic expression is found by means of operations (1÷4) executed on the subsets whose labels are primary terms. It is to be noted that:

- operation NOT is used in cases of words like "no", "not", "neither", "nobody", "never", etc.

- operation AND is used in cases of words "and", "but" and other words with similar meaning ("whereas", "while", etc.)

- operation OR is used in cases of words "or", "either ... or", "neither... nor" (in combination with NOT)

- operation POWER α is used in cases of linguistic hedges (for every hedge the value of α is fixed); for example "very" and "much" correspond to the operation CONCENTRATION ($\alpha = 2$), "little" and "slightly" - to the operation DELATION. ($\alpha = 0{,}5$) etc. (More detailed - see [1]).

DESCRIPTION OF GQS

The proposed general questionnaire system (GQS) uses the operations on fuzzy sets described in the previous chapter in order to provide quantitative estimates of the fuzzy answers received by the inquiry. The system overcomes a basic difficulty that arise when dealing with the fuzzy system theory apparatus - a determination of membership function values. Making use of the knowledge of questionnaire participators, GQS find estimates of those primary terms for which is expected to be used mostly in human answers. A general scheme of GQS is shown in figure 2.

1. The purpose of the system is to catch as completely as possible the basic tendency and nuances of public opinion of problems chosen beforehand by proceeding the needs of the control block (see fig. 1) in information.

2. By question-form making it also proceeded from the information necessary to the control block.

3. An auxiliary tour of inquiry is held simultaneously with the questionnaire realization. The purpose of this tour is to find some quantitative estimates of the primary terms which are expected to be used in the inquired man's answers.

The primary terms that have to be estimated are offered to a part of the questionnaire participants in the form of lists of admissible answers composed by the inquirer for each of the questions put in the auxiliary tour.

The participators are to evaluate the functions of membership of each objekt of the considered set to the fuzzy subsets that correspond to the primary terms. The number of participators that take part in the auxiliary tour is determined on the basis of some statistical considerations connected with the confidence intervals of the membership function estimates 2 .

4. The processing of the auxiliary tour's results consists merely in finding avarage values of primary terms estimates for each analysed objects.

5. The questionnaire (i.e. the basic tour of inquiry) is carried out with all participators. They are requested to write freely composed appropriate answers to the offered questions. Only a "direct" word-order of the sentence of the answers is required to be used, that is subject attribute - subject - predicate - predicative attribute (object, adverb).

6. The answers are processed in the following sequence:

i. The whole text of each answer is broken to simple clauses.

ii. Only the predicative attributes of the clauses are taken into account, since they bear the essential information about the attitude of the inquired ones to a considered question.

iii. The predicative attributes of each simple clause is taken up as a linguistic expression and is estimated by means of technique stated in the previous chapter.

iv. The same technique is applied to express a general estimate of the whole answer using the estimates of the simple clauses and opperations corresponding to the connectives between them (if there is no special connectives, the operation AND is applied).

7. By the analysis of each simple clause a separation of primary terms, linguistic hedges, negations, connectives is made. If the separated primary terms have not been estimated by the auxiliary tour of inquiry, their influence on the estimatesof the whole answer in which they take part is tested, i.e. a number of values representing the "unknown" primary term's estimate is generated and the corresponding values of the answers are found; if these values do not differ substantially, one of them is taken as an answer's estimate, otherwise the answer is considered as a non-processed and the primary term which influence was tested, has to be memorized.

8. After processing of all answers a decision for repetition of the auxiliary tour of inquiry or for making a list of new questions must be taken in dependence on the percentage of the processed answers (or on some other considerations connected with the qualitative features of the information accumulated as a result of the questionnaire realization.

On figure 3 a flow-chart representing the basic principles of the questionnaire processing is given.

Let us illustrate the functions of GQS by an example:

1. An auxiliary tour of inquiry is realized with University students in order to receive the estimates of the primary terms offered to them as answers to a question about the abilities (the talent) of the teachers u_1, u_2, u_3 and u_4 in reading their lectures to the students:

primary terms	estimates of for u_1	u_2	u_3	u_4
in a captivating way	0,7	0,7	0,4	0,3
interesting	0,8	0.9	0.1	0.9
quickly	0.6	0.6	0.3	0.5
comprehensibly	0.4	0.8	0.4	0.4
well	0.7	0.1	0.8	0.2
with erudition	0.9	0.4	0.8	0.7

Let us explain some operations on fuzzy subsets applied by the processing of the questionnaire.

The set $U = \{u_1, u_2, u_3, u_4\}$ is a set of the lecturers.
According to the table above, the grade of membership of each lecturer to the fuzzy subset U_4 (reading in a captivating way) and to the fuzzy subset U_9 (arousing interest) or the estimates of the primary terms "in a captivating way" and "interesting", respect. are:

"in a captivating way" $= U_4 = \{0.7|u_1, 0.7|u_2, 0.4|u_3, 0.3|u_4\}$

"interesting" $= U_9 = \{0.8|u_1, 0.9|u_2, 0.1|u_3, 0.9|u_4\}$

If it is necessary to find the estimates of terms:"not in a captivating way" and "not interesting", the application of operation (1') gives:

"not in a captivating way" $= \bar{U}_4 = \{0.3|u_1, 0.3|u_2, 0.6|u_3, 0.7|u_4\}$

"not interesting" $= \bar{U}_9 = \{0.2|u_1, 0.1|u_2, 0.9|u_3, 0.1|u_4\}$

The quantitative estimate of the expression that consists of the two primary terms connected with "and" , according to (3'), is

"in a captivating way and interesting" $= U_4 \wedge U_9 = \{0.7|u_1, 0.7|u_2, 0.1|u_3, 0.3|u_4\}$

The estimate of the expression with "or", according to (4'), is

"in a captivating way or interesting" $= U_4 \vee U_9 = \{0.8|u_1, 0.9|u_2, 0.4|u_3, 0.9|u_4\}$

Let us find the estimates of the

expressions "very much interesting" and "not very interesting and not in a captivating way". The application of operation POWER α with $\alpha = 2$ (according to (2')) gives:

"very much interesting" = very much (u_q)

$$= \text{very } (u_q)^2 = (u_q)^4$$

$$= \{0.4|u_1,\ 0.6|u_2,\ 0|u_3,\ 0.6|u_4\}$$

"not very interesting and not in a captivating way" = $(u_q)^2 \wedge u_f =$

$$= \{0.3|u_1,\ 0.2|u_2,\ 0.6|u_3,\ 0.2|u_4\}$$

/The highest grade of membership to the fuzzy subset "not very interesting and not in a captivating way" has the lecturer u_3 (0.6)/.

2. By the realization of the questionnaire, the inquired A has given the following answer to the question:"Does u_1 read his lectures good?" - "u_1 reads his lectures in a captivating way and very interesting. Well, I must say, he speaks a little quickly or not quite comprehensible, but it doesn't hinder him to be a good and exceptionally erudite lecturer". The corresponding estimates received by the processing of the answer (using the data for u_1 from table above) are:

in a captivating way = 0.7

very interesting = $(\text{interesting})^2 = 0.8^2 = 0.64$

a little quickly = $(\text{quickly})^{\alpha=0.5} = 0.6^{0.5} = 0.77$

not quite comprehensible = not $(\text{comprehensible})^{\alpha=2} = 1 - 0.4^2 = 0.84$

good = well = 0.7

exceptionally erudite = $(\text{erudite})^{\alpha=2.5} = 0.9^{2.5} = 0.75$

The estimates of the simple clauses are correspondingly:

first cluse: $\min(0.7, 0.64) = 0.64$

second clause: $\max(0.77, 0.84) = 0.84$

third clause: $\min(0.7, 0.75) = 0.70$

The estimate of the whole answer is

$$\min(0.64, 0.84, 0.70) = 0.64$$

0.64 is the grade of membership of the lecturer u_1 to the fuzzy subset "good lecturer".

Having on disposal the avarage estimates of all answers, GQS can offer decisions, coserning the career of u_1, u_2, u_3 and u_4 as lecturers, to the block of control (see fig.1).

GQS is realized on the computer MINSK 32. The software realization is in language of assembler; the details of the realization are not discussed here.

CONCLUSIONS

GQS was applied for finding the quantitative estimation of the attitude of the Sofia University students to the whole process of instruction (lectures, exercises, out-of-class work, how the students spend their free time, etc.). A merit of the system is its "ability" not to to be "deranged" when chancing upon fuzzy, pure human linguistic expressions (it is known that the functioning of all described formerly systems of artificial intelligence is paralysed when they come across such fuzzy terms). This makes us think that GQS is a step towards a complete artificial intelligence synthesis. As to the control of complex systems like groups of people, GQS could be a really useful and indispensable adviser to the block of control (governing). Some shortcomings of the system like unskilful handling of synonyms and others are a result of the salient features of the natural languages [3].

NOMENCLATURE

a_{nm}	-	quantitative estimate of A_{nm}
a_m	-	avarage quantitative estimate of the m-th question
A_{nm}	-	answer of the n-th inquired to the m-th question
K_{nm}	-	number of non-processing answers to the m-th question
M	-	total number of the questions offered by the questionnaire
m	-	number of a question
N	-	total number of the questionnaire participators
n	-	number of an inquired man
PT	-	primary terms

REFERENCES

1. Zadeh, L.,Fuzzy Systems, IEEE Trans. on Syst.,Man.Cybern., 1973, N1
2. Йетс Ф., Выборочный метод в переписах и обследованиях,М,1965
3. Бърнев П.,Добрев Д.,Статистични изследвания на българския език, Изв. Матем. инст.,1972

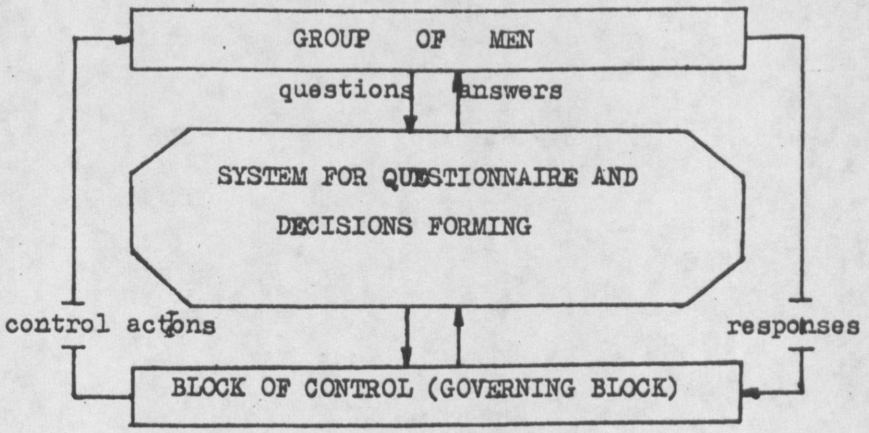
GROUP OF MEN
questions
answers
SYSTEM FOR QUESTIONNAIRE AND DECISIONS FORMING
control actions
responses
BLOCK OF CONTROL (GOVERNING BLOCK)

Fig.1

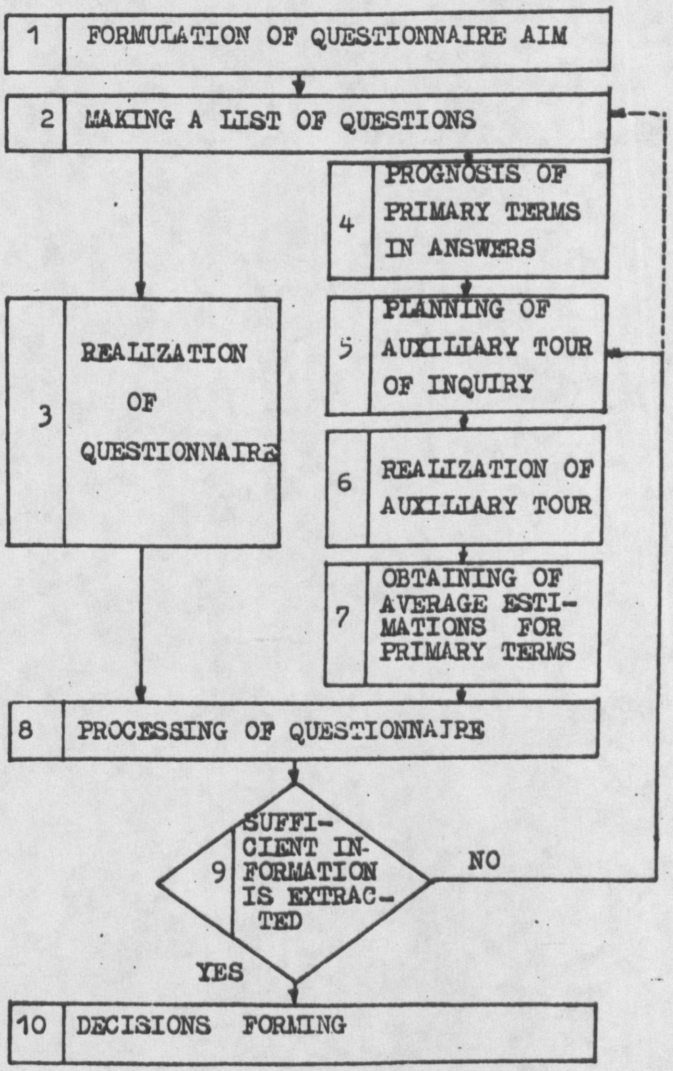
1 FORMULATION OF QUESTIONNAIRE AIM
2 MAKING A LIST OF QUESTIONS
3 REALIZATION OF QUESTIONNAIRE
4 PROGNOSIS OF PRIMARY TERMS IN ANSWERS
5 PLANNING OF AUXILIARY TOUR OF INQUIRY
6 REALIZATION OF AUXILIARY TOUR
7 OBTAINING OF AVERAGE ESTIMATIONS FOR PRIMARY TERMS
8 PROCESSING OF QUESTIONNAIRE
9 SUFFICIENT INFORMATION IS EXTRACTED
NO
YES
10 DECISIONS FORMING

Fig.2

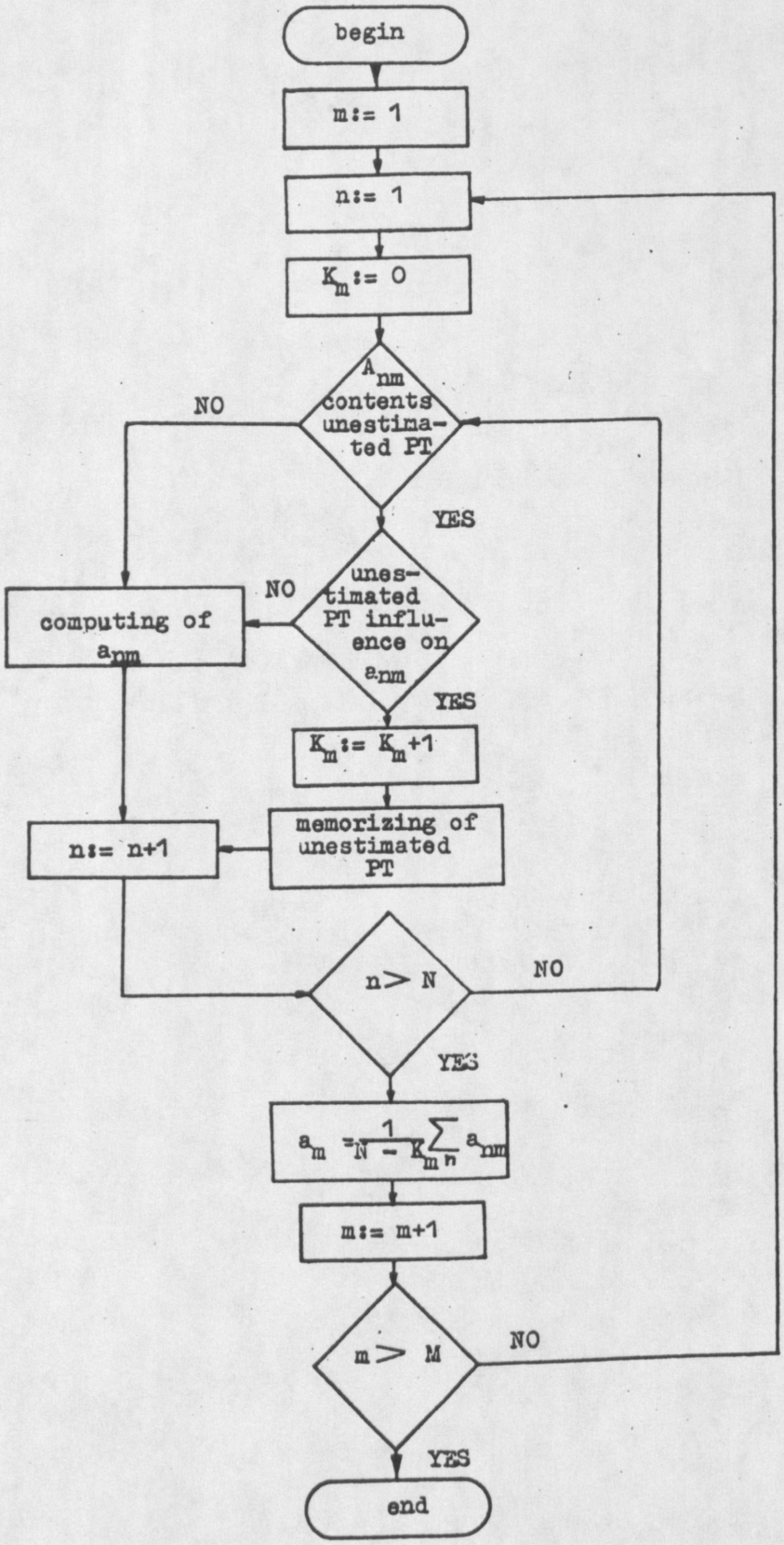
begin
m:= 1
n:= 1
K_m:= 0
A_{nm} contents unestimated PT
NO
YES
unestimated PT influence on a_{nm}
NO
computing of a_{nm}
YES
K_m:= K_m+1
memorizing of unestimated PT
n:= n+1
n > N
NO
YES
$a_m = \frac{1}{N - K_m} \sum_n a_{nm}$
m:= m+1
m > M
NO
YES
end

FIG. 3

(84.) Barnev P., Dimitrov V., Stanchev P., "Fuzzy System Approach to Decision Making Based on Public Opinion Investigation through Questionnaires", **Proc.** ***IFAC Symp. of Stochastic Control,*** Budapest 1974, 637-642.