M. STOICA
I. M. STANCU—MINASIAN

Computation of Statistical Means with the help of the Theory of Graphs

A method is presented similar to that used in the computation of a medium value by way of the least squares method, for the computation of statistical means with the help of the theory of graphs. The method worked out by the authors facilitates the computation of statistical means whose values thus obtained tend approach those of the means of means.

PRESENTATION OF THE PROBLEM

A big number of experiments grouped according to certain criteria are made in the practice of statistical measurements. Suppose that we have n groups of experiments:

$$M_1 = \{a_1^{(1)},\ a_2^{(1)}, \dots,\ a_{p_1}^{(1)}\}$$
$$M_2 = \{a_1^{(2)},\ a_2^{(2)}, \dots,\ a_{p_2}^{(2)}\}$$
$$\dots\dots\dots\dots\dots$$
$$M_n = \{a_1^{(n)}, \dots,\ a_{p_n}^{(n)}\}$$

For a group of experiments M_i, p_i experiments were made and the data $a_1^{(i)}$ ... $a_{p_i}^{(i)}$ were obtained. The set $M = M_1 \cup M_2 \cup \dots \cup M_n$ of these multitutes of data is to be processed. Situations of this kind frequently appear in a series of economic problems. We confine ourselves to some of them.

1) Suppose that we want to determine the necessary time for carrying out the processing operation of a machine part. For this, we make chronometric observations considering groups of workers in various sections where

the respective operation is carried out. In section S_i, p_i workers are timed and the results $t_1^i, \ldots t_{p_i}^i$ are obtained. In most cases, the workers' aptitudes are different, which means that the data obtained will not be identical. The question is how to choose from each group of data a representative value which should be taken into consideration in computing the mean.

2) The determination of the average rate of increase of the production of some enterprises in a branch of the national economy, for example, the machine-building branch, within a five-year plan, necessitates the processing of the data collected from several enterprises. In this case, let us admit that the enterprises $I_1, \ldots, I_s$ are analysed and for each enterprise I_k $(k = 1, \ldots, s)$, the rates $r_1^{(k)}, \ldots, r_{t_s}^{(k)}$ are considered in various periods of time. The aim pursued is the same as in the first example.

3) p equations are given, the coeficients of which result from practical measurements:

$$a_1^{(1)}x^{n_1} + a_2^{(1)}x^{n_1-1} + \ldots + a_{n_1+1}^{(1)} = 0$$

$$a_1^{(2)}x^{n_2} + a_2^{(2)}x^{n_2-1} + \ldots + a_{n_2+2}^{(2)} = 0$$

$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$a_1^{(p)}x^{n_p} + a_2^{(p)}x^{n_p-1} + \ldots + a_{n_p+1}^{(p)} = 0$$

about which it is known that they have q common roots $(q \leqslant min(n_1, n_2, \ldots, n_p)$. Due to the fact that errors might occur in the determination of the coeficients a_j^i, by solving these equations, it might happen that not all the q common roots and even none should be obtained. In this case, q roots close to the common ones should be found.

4) In solving the problems of analysis of the critical path, it is necessary to know most exactly the durations of the activities. In the activities which are carried out for the first time, simple appreciations are made, taking three durations into consideration: the pessimistic, optimum and the likely durations. For the activities which have already been carried out, data from the the previously carried out works are used for a most realistic estimation of the average durations. It is a problem having the same character as the above mentioned ones.

SOLVING OF THE PROBLEM

Consider n sets of real numbers,

$$M_1 = \{a_1^1,\ a_2^1,\ a_3^1,\ \ldots\ ,\ a_{p_1}^1\}$$

$$M_2 = \{a_1^2,\ a_2^2,\ a_3^2,\ \ldots\ ,\ a_{p_2}^2\}$$

$$\ldots\ldots\ldots\ldots\ldots\ldots$$

$$M_n = \{a_1^n,\ a_2^n,\ a_3^n,\ \ldots\ ,\ a_{p_n}^n\}$$

in which we suppose that for each set M_i, the elements are written in a growing order, that is:

$$a_i^{(k)} \leqslant a_j^{(k)} \qquad k = 1, \ldots, n \ (i \leqslant j)$$

We may also suppose that the sets are arranged according to the elements contained by them, which means that $p_i \geqslant p_j$ $(i \leqslant j)$. Both hypotheses do not restrict the generality of the problem.

We propose to find a $\bar{x}$ number which minimizes the expression:

$$\sum_{k=1}^{n} (\min_{i} | x - a_i^k |)^2$$

To find the number $\bar{x}$, we shall seek in each set M_i an element $a_{j_i}^i$ which should be taken into consideration in calculating the mean.

In this case, the average value $\bar{x}$ will be:

$$\bar{x} = \frac{\sum_{i=1}^{n} a_{j_i}^i}{n}$$

To solve this problem, we shall use methods of the theory of graphs.

To this purpose, a graph is build, as follows: a point in the plan $P_{p_i}^i$, is associated to each value $a_{p_i}^i$, all the points of a set M_i being placed on the same level (figure 1). Each top $P_{p_i}^i$ on the level i is united with all tops situated on the level $i + 1$, of an arc $(P_{p_i}^i, P_{q_{i+1}}^{i+1})$ directed from $P_{p_i}^i$ to $P_{q_{i+1}}^{i+1}$, giving it a value equal to the square of the module of the difference of the numbers corresponding to the points, that is

$$| a_{p_i}^i - a_{q_{i+1}}^{i+1} |^2$$

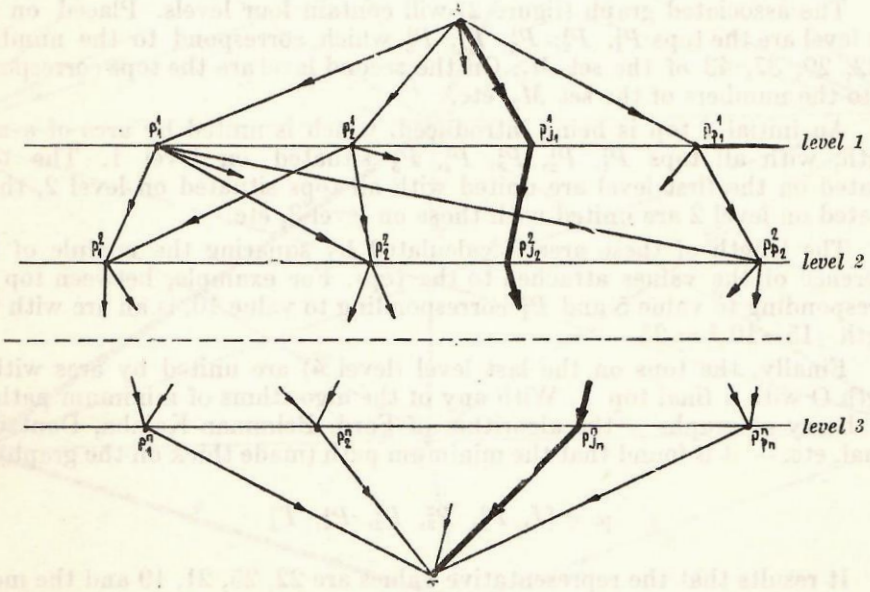

Fig. 1

An initial top I is being introduced, which unites with all the tops P_j^1 situated on the first level, through arcs 0 long and with a final top T to which come arcs from the tops situated on the last level, arcs having a length of zero.

In this way, the finding within each group of data of the sort to be used in the computation of the mean, is limited to the finding of the minimum path in the associated graph.

Consider $\mu = (I, P_{j_1}^1, ..., P_{j_i}^i, ..., P_{j_n}^n, T)$ — the minimum path found. The representative values will be $a_{j_i}^1, ... a_{j_n}^n$, and the average value

$$\bar{x} = \frac{\sum_{i=1}^{n} a_{j_i}^i}{n}$$

NUMERICAL EXAMPLE

To illustrate the above theoretical considerations, we suppose that four groups of experiments with the values:

$M_1 = \{ 5, 22, 29, 37, 43\}$
$M_2 = \{10, 16, 25, 30\}$
$M_3 = \{20, 21, 59\}$
$M_4 = \{19, 30, 40\}$

were carried out.

The associated graph (figure 2) will contain four levels. Placed on the first level are the tops P_1^1, P_2^1, P_3^1, P_4^1, P_5^1, which correspond to the numbers 5, 22, 29, 37, 43 of the set M_1. On the second level are the tops corresponding to the numbers of the set M_2, etc.

An initial I top is being introduced, which is united by arcs of a zero length, with all tops P_1^1, P_2^1, P_3^1, P_4^1, P_5^1 situated on level 1. The tops situated on the first level are united with all tops situated on level 2, those situated on level 2 are united with those on level 3, etc.

The length of these arcs is calculated by squaring the module of the difference of the values attached to the tops. For example, between top P_1^1 corresponding to value 5 and P_1^2 corresponding to value 10, is an arc with the length $|15-10|^2 = 25$

Finally, the tops on the last level (level 4) are united by arcs with a length O with a final top T. With any of the algorithms of minimum path of the theory of graphs — the algorithm of Ford, Belmman-Kalaba, Dantzing, Arinal, etc. — it is found that the minimum path (made thick on the graph) is:

$$\mu = [I, P_2^1, P_3^2, P_2^3, P_1^4, T]$$

It results that the representative values are 22, 25, 21, 19 and the mean is 21. 75.

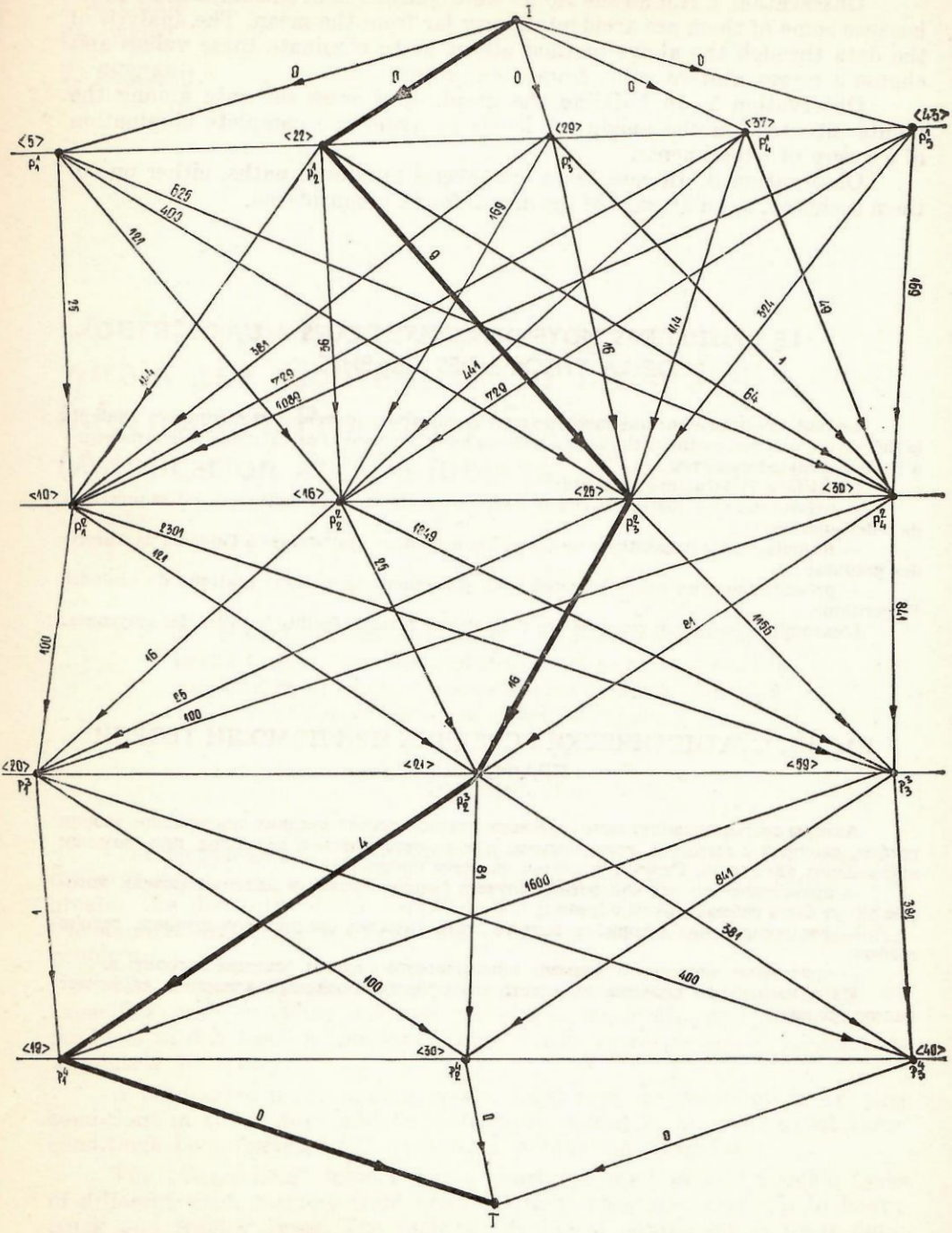
I
T
<5>
<22>
<29>
<37>
<43>
<10>
<16>
<25>
<30>
<20>
<21>
<59>
<19>
<30>
<40>
0
625
400
121
169
25
381
36
729
1089
441
16
144
324
67
1
64
2301
1249
100
81
1156
841
4
1600
400

Fig. 2

Observation 1. Not all the values were considered in calculating the mean because some of them are accidental, very far from the mean. The analysis of the data through the above method allows us to eliminate these values and choose a representative value from each group.

Observation 2. In building the graph, arcs were set only among the points situated on the neighbour levels to avoid the complete elimination of a group of experiments.

Observation 3. In case there are several minimum paths, either one of them is chosen, or an average of the means found is considered.

LE CALCUL DES MOYENNES STATISTIQUES A L'AIDE DE LA THEORIE DES GRAPHES

Les auteurs présentent une méthode pour le calcul des moyennes statistiques à l'aide de la théorie des graphes, méthode similaire à celle qu'on utilise pour le calcul d'une valeur moyenne à l'aide des moindres carrés.

L'article a la structure suivante:

— présentation de quatre cas concrets (économiques et mathématiques) qui se prêtent à de telles solutions;

— formulation de l'algorithme de calcul des moyennes statistiques à l'aide de la théorie des graphes;

— présentation d'un exemple numérique, concernant la manière pratique de résoudre l'algorithme.

L'exemple présenté fait ressortir que l'algorithme formulé facilite le calcul des moyennes.

РАСЧЕТ СТАТИСТИЧЕСКИХ СРЕДНИХ ПРИ ПОМОЩИ ТЕОРИИ ГРАФОВ

Авторы статьи приводят метод расчета статистических средних при помощи теории графов, сходный с методом, применяемым при расчете средней величины при помощи наименьших квадратов. Структура статьи является следующей:

— приведение четырех конкретных случаев (экономических и математических), которые могут быть решены таким образом;

— формулирование алгоритма расчета статистических средних при помощи теории графов;

— приведение численного примера практического способа решения алгоритма.

Из приведенного примера вытекает, что сформулированный алгоритм облегчает расчет средних.