QUANTITATIVE ANALYSIS OF SOME TENDENCIES IN THE DEVELOPMENT OF OPERATIONS RESEARCH AND MATHEMATICAL ECONOMICS

L. S. Kozachkov, L. A. Khursin, and N. Z. Shor

Kibernetika, Vol. 3, No. 5, pp. 102–108, 1967

UDC 002:519.8

One of the most important problems of the provision of information for scientific investigations is the study of the tendencies of development of individual branches of science. A fundamental method of determining the tendencies of scientific progress at the present time is the direct study of scientific literature and the preparation of surveys of the corresponding areas of science. But this method, which may be termed qualitative or subjective, must be supplemented by a quantitative method of investigating the development trends of some field of knowledge as a whole and of its individual areas. The theory of scientific information and scientific knowledge must equip scientists with quantitative methods of studying the development tendencies of science.

However, even the theory of scientific information and scientific knowledge has really taken only the first steps toward the development of scientifically based quantitative methods of estimating scientific and technological progress. At present descriptive methods continue to dominate in these sciences, and the principal apparatus of investigation is statistics. Very rarely is use made of modern methods of mathematical simulation enabling complex laws to be described dynamically.

Here, as we pointed out by Glushkov [1], it is a question not of physical but of informational simulation of the complex internal properties of the system investigated; moreover this reveals the possibilities for the extensive utilization of that universal instrument of dynamic information simulation, the digital computer.

Among the papers devoted to the study of the laws of information flows, it is worthwhile to note (for their methodology and use of the elements of modern mathematics) the investigations of E. Garfield, D. J. Price, and in our country G. M. Dobrov, V. V. Nalimov, and their collaborators.

In this paper an attempt is made to justify the possibility of using quantitative indexes of information flows together with a quantitative analysis of the state of scientific trends, for determining the development tendencies of science. For this purpose we have selected a field of applied cybernetics, operations research and mathematical economics, and as the method of qualitative estimation of the state of science the so-called law of the "scatter" of information is used [2].

In the monograph [3] the following description of the law is given. If the set of all the published items devoted to some question is taken as unity, then the special periodicals on this area of knowledge, whose number is as a rule not great, contain only a third of these published items. The second third of the articles on this branch of knowledge are published in a much more extensive set of journals which are topically related but nevertheless in another sector. And, finally, the last third of published items is scattered over an enormous number of periodicals of the most varied type.

In analytic form this can be written as follows:

$$T_x : T_{2x} : T_{3x} = 1 : n : n^2, \qquad (1)$$

where T_X, T_{2X}, T_{3X}, are the number of periodical issues containing respectively x, 2x, 3x papers on the given subject. The number n depends on x and has a different value in each branch of knowledge. If it is assumed

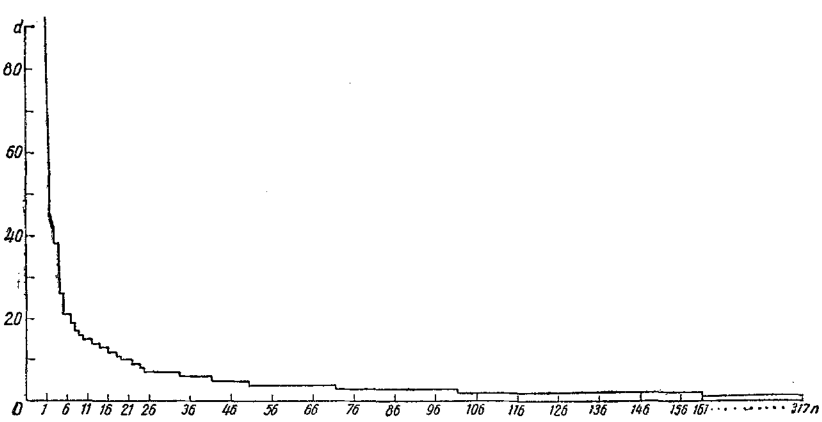

Fig. 1. Distribution of published items by periodical issues. d is the number of published items in a periodical issue; n is the order of a periodical issue in a series in decreasing order of the number of published items.

that the whole set of periodical issues consists of three zones [A) specialist sources, B) related sources, C) the remainder] and it is assumed that each zone contains one third of all the information, then the total quantity of periodical issues T_0 containing all the information on the given question is easily calculated as

$$A:(A+B):(A+B+C)\approx 1:n:n^2. \qquad (2)$$

Hence,

$$T_0=\frac{(A+B)^2}{A}. \qquad (3)$$

The predicted values of T_0 have been compared with statistics on the following branches of cybernetics: the theory of automata, the theory of mathematical machines, operations research, and mathematical economics. Analysis of the statistics (for the preliminary analysis use was made of the abstracts journals Kibernetika, Matematika, and Ekonomika promyshlennosti for the years 1956–1966) did not confirm the correctness of the quantitative parameters of the scattering law on the division of all the information into three zones each containing a third of the total number of published items. For example, according to the model considered we obtained that the abstracts journals lose about 1000 periodical sources containing information on the theory of automata, though in reality the loss of this literature is described

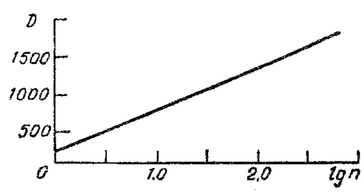

Fig. 2. Distribution of papers in periodical sources according to M. G. Kendall's data.

by a much more optimistic figure. However, in the field of operations research, if we base ourselves on the parameters of the scattering law indicated above, it is possible to conclude that the abstracts journals encompass 100% of all the periodical issues containing papers on this subject. In actual fact, in accordance with the data of the abstracts journals for the years 1965–1966, 10 of the main sources already contained one third of the published items (682), the remaining sources detected contained two thirds of the information (1041 papers). But the conclusion that the abstracts journals completely cover all the literature on operations research does not agree at all either with the data of the abstracts journals for other years, or with the Science Citation Index of E. Garfield for the year 1965, which made it possible for V. V. Nalimova in studying references to the papers of R. Bellman to discover periodical issues not taken into account in our statistics.

Figure 1 shows the distribution of information by periodicals for the year 1962, in the order of decrease of the number of papers contained in them. The distribution has a similar form in other years also.

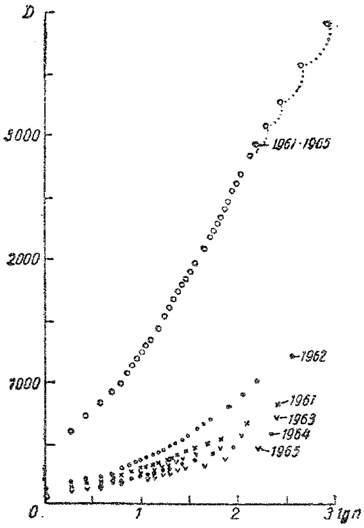

Fig. 3. Distribution of articles in periodical sources for the years 1961–1965.

What has been said above gives a basis for concluding that the "information scattering" law of Bradford and Vickery, derived from statistics relating to different fields of knowledge, does not, without additional investigations, permit this important objective law of information flows to be used for the purpose of predicting the development of science.

This law has been studied by the well-known specialist in operations research. Kendall [4], in examining the applicability of the law of scattering in the field of operations research. It appears that if the law of scattering is represented graphically, by laying off along the horizontal axis the logarithm of the increasing number of the sources ln n, arranged in decreasing order of the number of papers contained in them, and along the vertical axis the grand total of published items D, there is a straight-line dependence which Kendall generalized in the following words: "The linearity is excellent and in fact better than in Bradford's own examples" [4].

Indeed, this linearity is also confirmed by the processing of statistical material obtained by us by the selection of items on operational research and mathematical economics from the abstracts journals for the years 1961–1965.

The graphical representation of the "scatter" in the coordinates (D, ln n) gives a straight line in any of the years considered. Only the slope of the straight line changes.

However, the straight lines given in Fig. 3 and the values of the angles corresponding to them give only a visual impression of the "scattering" process, since the construction is carried out on an arbitrary scale (the same as in the similar graphs of Bradford and Kendall), and the approximation of the points by the straight line was made without using the necessary methods of mathematical statistics.

Approximating the set of points by a straight line by the method of least squares, we can write the equation of the straight line in the following form:

$$D(n) = a + \operatorname{tg}\varphi \ln n, \quad (4)$$

where a is the value of the number of papers in the first source obtained by approximating the set of points by a straight line by the method of least squares; φ is the angle of slope of the straight line to the horizontal axis.

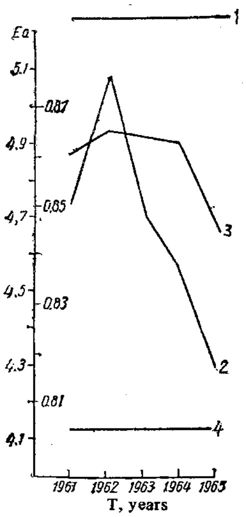

Fig. 4. Variation of the entropy of a system of papers distributed over sources: 1) E_a cumulative (1961–1965); 2) E_d; 3) E_{rel}; 4) E_{rel} cumulative (1961–1965).

It must again be emphasized that in the general case the parameter a does not coincide numerically with the number of papers in the real first source of an ordered series of periodical issues; a is a parameter of the mathematical model adequately reflecting the law of the distribution of papers over the whole set of information sources considered.

The number of papers in each source d in the ordered series of sources is a function of the number of the source; we will denote it by $f(n)$.

The total number of papers is

$$D(k) = \sum_{n=1}^{k} f(n)\, dn. \quad (5)$$

For convenience of discussion we replace the discrete argument function $f(n)$ by the continuous argument function $\bar{f}(n)$, and the operation of summation by integration, so that for sufficiently great n

$$f(n) \approx \bar{f}(n) \quad (6)$$

and

$$D(k) = \int_{1}^{k} \bar{f}(n)\, dn. \quad (7)$$

Differentiating equation (4), we obtain

$$\frac{dD}{dn} = \frac{\operatorname{tg}\varphi}{n}. \quad (8)$$

Since it follows from the expression (7) that

$$\frac{dD}{dn} = \bar{f}(n) = f(n), \quad (9)$$

then, taking into consideration the relation (8), we have

$$d \cdot n = \operatorname{tg}\varphi. \quad (10)$$

Equation (10) considered above is nothing but the equation of the equilateral hyperbola (in the coordinates d, n), obtained by transforming the straight line (4) (in the coordinates D, ln n).

Hence the following proposition can be formulated: in an ordered set of information sources the number of publications they contain decreases in accordance with an equilateral hyperbolic law.

From the nature of the hyperbola describing the distribution of articles in an ordered series of sources there follows an important practical result: on the average, the statistical product of the number of sources by the number of articles on a given branch of knowledge contained in them, in a definite interval of time, is a constant quantity equal to tg φ. This result, which follows from the consideration of our theoretical model, is confirmed both by our statistics on operations research and by the statistical data published by other authors.

Therefore, the study of the question of the distribution of scientific information in periodical issues carried out by various authors during the last 20 years led to the appearance of several models describing this process (S. C. Bradford, B. C. Vickery, M. G. Kendall, and others). All of the proposed models of the distribution of information in sources had as their object the illustration of the phenomenon of the "scatter" of papers. Hence this law of the distribution of papers over sources was called the law of the scattering of information. In reality, however, this important law of the distribution of papers describes rather the concentration of information.

Indeed the whole accumulation of statistics on the distribution in periodical sources of information on some field of knowledge testifies to the fact that in the given subject are concentrated all the fundamental investigations in a comparatively restricted number of periodical sources. Kendall also turned his attention to the fact that the first 18 periodical issues, that is, 5% of all the observed sources (370), contained about 50% of all the papers on operations research and statistics.

This is also confirmed by our statistics in which 50% of the information is contained in 37 sources which also compose 5% of the total number (686) of the periodicals we found, which publish information on operations research and mathematical economics.

This gives reason to assume that our mathematical model of the distribution of information in accordance with an equilateral hyperbola actually describes a law of information concentration, and the hyperbola itself may be called the hyperbola of information concentration. In our opinion this result is of fundamental importance and reflects a deeper feature of the distribution of information than formally follows from the model of the "scatter" of information considered previously. The impression of the scattering of information of authors who are engaged on the study of a given question, is based on the study of a whole large-scale field of knowledge, geophysics (S. C. Bradford), electrical engineering (D. U. Taplow), operations research and statistics (M. G. Kendall). However, any field of knowledge consists of individual sections differing from one another both in the methods of investigation and in the nature of the applications in other fields of science. In this respect there is nothing unique about the field of cybernetics we are studying, i.e., operations research and mathematical economics. It is obvious that the more extensive the field of knowledge of the science considered, the greater the "scatter" to be expected.

For example, cybernetics and its applications (here operations research is also included) will be "scattered" in an enormous set of issues, whose number is difficult to predict, and, conversely, information on a narrower sector of science is concentrated in an extremely restricted number of issues.

The division of the field of science considered into sectors, and the attachment of these sectors to zones where there is the greatest probability of the concentration of information on each of the sections of the subject, requires the combination of quantitative analysis of the periodical issues with qualitative (content) analysis.

In order to determine the nature of the distribution of the information in each of the zones we turn to Figs. 2, 3. The straight lines shown in the diagrams represent the distribution of information over the whole set of issues and over the whole of the subject considered. As we have seen, these straight lines correspond to the hyperbola of information concentration in the coordinates (d, n).

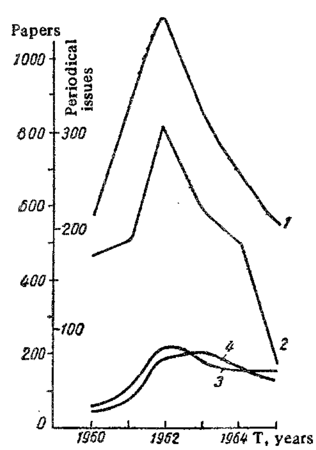

Fig. 5. Growth of papers and periodical issues. 1) Foreign papers; 2) foreign issues; 3) Russian issues; 4) Russian papers.

However, for a more specific description of the information concentration in each zone, and consequently, for purposes of probabilistic prediction of the places where papers on the individual sectors of the field of science can be found, it is no longer possible to remain within the framework of the model described above. In order to describe the information distribution process in each of the zones, we assume that for any source in the given zone, the quantity d = = b = const (b is the number of papers in the source).

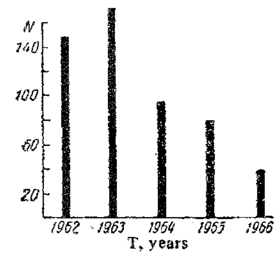

Fig. 6. Number of new sources.

In this case the cumulative number of papers in the zone is given by

$$D_3(n) = \int_{n_k}^{n} b\,dm = b\,(n - n_k), \tag{11}$$

where n_k is the ordinal number of the source before the beginning of the zone, and n is the running ordinal number of the sources of the zone considered.

Since

$$n = e^{\ln n}, \tag{12}$$

equation (11) in the coordinates (D, ln n) is represented in the form

$$D_3 = b\,(e^{\ln n} - e^{\ln n_k}). \tag{13}$$

This is an exponential equation in the assumed coordinate system (D, ln n). This implies that if the papers are uniformly distributed over sources, the relation between the cumulative value of the papers and the logarithm of the ordinal number of a source, is not linear but exponential. Therefore, in the general case to a second approximation the distribution of papers in a system of ordered sources may be represented graphically as in Fig. 3.

In the general case, a system of periodical issues separates into a series of zones having different mathematical descriptions. Issues with a comparatively large number of articles on a given problem we call the "core." For the "core" there is a small probability of coincidence of the number of papers in two different sources, and the information distribution is well enough described by a straight line in the coordinates (D, ln n). Issues with a comparatively small number of papers (1 to 10) clearly fall into a series of zones with an identical number of papers, each of which is described by its segment of the exponential. In this connection, the "tail" of the distribution of papers is not well approximated by a straight line.

The laws of information flows reflect the laws of development of the science itself and of its individual fields. Hence by studying the laws on information distribution we can forecast some tendencies of scientific research. However, it does not appear to be possible to use for this purpose the results of previous investigations in this field, since the authors, in studying the "scatter" of information, have restricted themselves to a static description of this process, that is, they have not analyzed the nature of the distribution of papers in sources as a function of time. The study of the dynamics of this phenomenon is extremely important for an understanding of the process itself, and also for its prediction.

In order to construct a dynamic model of the distribution of information over sources, it is necessary to have a parameter which will give a sufficiently good description of the system as a whole. For this purpose, by analogy with the concept of entropy in Shannon's theory of information, we introduce the concept of the entropy of a system of distribution of information over sources, described by the relation

$$E = -\sum_{i=1}^{N} \frac{d_i}{D_N} \ln \frac{d_i}{D_N}, \quad (14)$$

where N is the number of sources containing D_N papers, d_i is the number of papers in the i-th source.

From this, N is the number of states of the system; d_i/D_N is the probability of the concentration of information in each of the states. The entropy of the system of sources is calculated in the same way, as the sum of the products of the probabilities of information concentration and the logarithms of these probabilities. The entropy depends on the number of sources N, and on the probability of the concentration of information in each source. The entropy of a system equals zero when all the papers are concentrated in one source, it increases with the growth of N and attains a maximal value when the number of papers in the sources is equiprobable. In this case

$$E_{max} = \ln N, \quad (15)$$

that is, the entropy of a system of sources where the occurrences of papers are equiprobable, equals the logarithm of the number N of sources.

In order to estimate the dynamics of a system of papers during the years 1961–1965 we construct a graph of the variation of the absolute and relative entropy (Fig. 4).

It is obvious from the diagram that the absolute entropy of the system of information sources in operations research and mathematical economics attained its greatest value in 1962, and then fell from year to year. The decrease of entropy after 1962 describes the process of concentration of the information into the basic core of the system of information sources.

It is natural to suppose that the fact of a steady fall of entropy in the system of papers reflects some essential feature of the development of the given branch of knowledge.

The majority of the sectors of the mathematical apparatus of cybernetics—the theory of probability, the theory of information, mathematical logic, and others, existed before cybernetics, and the results obtained in these fields were published (and frequently continue to be published) in specialized mathematical journals. As cybernetics became a more mature discipline, specialized journals devoted to its problems began to appear.

Later cybernetics began successfully to penetrate into other fields of knowledge, primarily into economics, medicine, biology, and the like. An enormous number of papers connected with the applications of cybernetics were published in the journals of those fields where these methods were applied. This process of the outward growth of cybernetics has not stopped and it may be that the history of science and technology contains no other example of such a rapid propagation of the methods of one science into others. It would appear that in these conditions it must be supposed that the number of papers and periodical sources devoted to the given subject must continue to grow, and consequently, a continuous increase in the entropy of the system can be expected. However, as we have seen, this assumption is not confirmed. Moreover, the fall of entropy is primarily caused by a decrease in the number of periodical issues publishing papers on the subject considered. This is illustrated in Fig. 5 by the curves of the variation of the number of sources and the number of papers for the years 1961–1965.

Attention has been drawn to the fact that the maximal number of sources and abstracts on foreign documents arrived in 1962, and on Russian ones in 1963, the number of papers in foreign sources in 1965 had fallen to almost a third of the 1962 number, and Russian papers to almost two-thirds.

There is no reason to suppose that this data testifies to a decrease in the quality of operation of the corresponding abstracts journals. Rather, this is an objective law connected with the variation of the character of investigations in this field of knowledge. The confirmation of this thought, and also the "rehabilitation" of the abstracts journal Kibernetika may be provided by our observation of the growth of a new

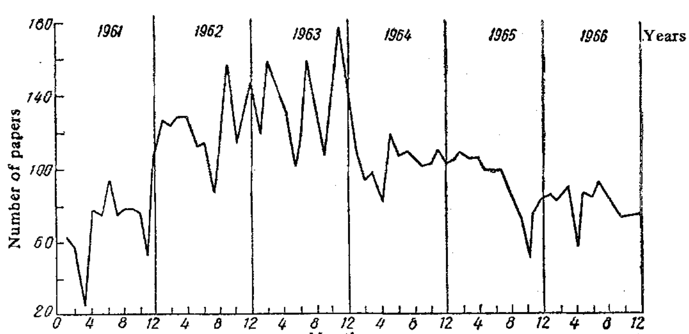

Fig. 7. Variation of the number of papers.

sources in each of the years in comparison with the preceding, which is reflected in the abstract journals (Fig. 6).

An analysis by content of these new sources testifies to the continuing penetration of the theory of optimal solutions into new regions of science and technology, although this process is not proceeding so rapidly now as four or five years ago, as is confirmed by the fall in the total number of papers.

Here there is an obvious contradiction with the so-called law of the exponential growth of papers. It must be pointed out that the law of the experimental growth in papers is mainly confirmed by statistics reflecting the state of major groups of sciences: natural science, mathematics, physics and so on. But the increase of publications devoted to any one field of science or technology often, as in our case, does not obey the exponential law. Attention was drawn to this fact by D. J. Price [5] and V. V. Nalimov [6], when they proposed that in certain circumstances the exponential law may become a logistic curve. However, the "damping" of the curve of growth of a science or its transition into a logistic curve still does not mean the "damping" of the given branch of knowledge, and can only testify to the detachment of a new science, the variation of the general direction of research and so on. An interesting feature of the passage of the exponential growth curve of a science into a logistic curve was mentioned by D. J. Price: "... obviously an exponential growth resents the idea of a bend if the former continues for a long time. Before the growth reaches the mean point it begins to jump and twist, to change, like mischievous spirits, its shape and features to avoid this terrible ceiling. Or, speaking in less anthropomorphic analogs, here a cybernetic hunting effect is established and the curve experiences sharp fluctuations.." [4].

This phenomenon is illustrated by the diagram of variation of the number of papers on operations research and mathematical economics, constructed by choosing a unit of time equal to one month, for the years 1961–1966 (Fig. 7) (from the data of abstracts journals).

It is obvious from the diagram that the "cybernetic hunting effect" and the "jumps" like "mischievous spirits" relate mainly to the years 1962–1963, after which a comparatively smooth fall of the curve is observed.

Therefore, qualitative analysis of the tendencies of the development of operations research and mathematical economics for the years 1961–1966, indicates the following: 1) the greatest absolute and relative entropy of the system of papers distributed among sources belongs to 1962, and then the value of the entropy falls; 2) the number of papers in periodical sources has a tendency to decrease beginning with 1962; 3) the years 1962–1963 are the end of some essential stage in the development of the science, which corresponds to the time of transition of the exponential into a logistic curve.

All these quantitative indications are only important symptoms of internal processes taking place in the science itself, and offer the possibility of determining the boundaries of the individual stages of the development of the science, of predicting its further tendencies, periods of transition from research and theoretical investigations to extensive applications, of determining the probable location in the system of issues of papers on fundamental investigations, various applications and the like. An analysis of papers in terms of content makes it possible to check the validity of conclusions based on the quantitative methods of studying processes, and also to explain the essence of the laws in the development of the science, which is adequately manifested in its quantitative characteristics.

The field of knowledge combined in our sector of cybernetics, operations research and mathematical economics, has a history of many years, but began to be intensively developed during the last twenty years.

Although the first papers of Academician L. V. Kantorovich on linear programming relate to the year 1939, they were not appropriately developed in the following years and were rediscovered in 1945–1947 by the American mathematician Dantzig. The fundamental work on the theory of games by J. von Neumann and Morgenstern was published shortly before this in 1944. The simple method developed and tested by Dantzig turned out to be an unexpectedly efficient procedure even to the author himself. At the same time high-speed computers appeared which facilitated the rapid development of methods of solving linear programming problems and their practical application. At the beginning of the fifties work in the United States on operations research and the application of mathematical methods in economics became so animated that a society was formed on operations research with its journal "Operational Research" founded in 1952.

In the course of the years 1950–1960 the theories and methods of convex programming were developed and brought to great perfection, in the years 1955–1960 dynamic programming and the theory of optimal processes, the theory of statistical decisions, 1960–1965 decomposition schemes for the solution of problems with a block structure, integer-valued programming, Gomory's method, the sequential analysis of variants, the method of "branches and boundaries." A great impulse to the development of scheduling theory and discrete programming generally was given by the development of network methods of planning and control begun in 1957. In recent years there has been a rapid growth in papers on stochastic programming, random search methods, the simulation of large systems, optimization problems on graphs and others.

In the Soviet Union work on the application of mathematical methods in economics intensified after the year 1955. At first (up to 1960) it was studied in a small number of groups under the direction of Academicians V. S. Nemchinov and L. V. Kantorovich in Moscow and Leningrad. The beginning of 1960 saw a tempestuous growth of the number of groups and specialists working in the field of the applications of mathematical methods in economics and operations research. Here must be mentioned the formation of important centers on these problems in Kiev (Institute of Cybernetics AS UkrSSR, Computer Center Gosplan UkrSSR), in Novosibirsk (Institute of Mathematics Siberian Division AS USSR), in Moscow (TsEMI, Computer Center Gosplan USSR) and in other cities. Therefore in the USSR and abroad work on operations research and the application of mathematical methods in economics is being developed both in breadth and in depth. What can explain the greater concentration of papers and the general decrease in references to periodical issues? It seems to us that this testifies to a certain maturity of this branch of knowledge.

Concentration is facilitated by the creation in sufficient quantity of specialized periodical issues, and also the appearance of fundamental monographs which causes a reduction of papers on secondary questions and technical details of the application of generally known methods, which were frequently encountered in the journals of other branches of knowledge. As regards the decrease of the total number of papers in periodical issues, this is overshadowed by the sharp increase in the output of monographs, various collections, and other nonperiodical editions.

These preliminary conclusions on the development tendencies of operations research are based on the supposition that the distribution of publications considered in time reflects some essential regularity of the science itself. "School arithmetic," for example, is "scattered" in all the journals not because there does not exist a specific journal on this subject, but because "school arithmetic" has ceased to exist as a living science and the distribution of its papers in the literature has attained maximal entropy. And, conversely, the history of science offers a large number of examples where some trend, once conceived in one scientific establishment, is continued only in papers of the printed journal of this establishment.

Between these extreme cases there is a set of states which are reflected in the real distributions. However, there is no "scatter" of information in any of the states of the system of publications.

The authors thank A. A. Patiokha and T. P. Sagaidak for selecting the statistical material and sorting it on punched cards, and also S. A. Kobtseva who helped in processing the material on the computer.

REFERENCES

1. V. M. Glushkov, "The epistemological nature of information simulation," Voprosy filosofii, no. 10, Moscow, 1963.

2. Handbook of Special Librarianship and Information Work (2nd ed.), W. Ashworth, gen. ed., London, Aslib, 1962.

3. A. I. Mikhailov, A. I. Chernyi, and R. S. Gilyarevskii, Fundamentals of Scientific Information [in Russian], izd-vo "Nauka", Moscow, 1965.

4. M. G. Kendall, The Bibliography of Operational Research, Operational Research Quarterly, vol. 11, numbers 1/2, March/June 1962.

5. D. J. Price, Little Science, Big Science [Russian translation], collection: The Science of Science, izd-vo "Progress", Moscow, 1966.

6. V. V. Nalimov, "Quantitative methods of investigation of the process of the development of science," Voprosy filosofii, no. 12, Moscow, 1966.

7. G. M. Dobrov, Science about Science. Introduction to General Scientific Knowledge [in Russian], izd-vo "Naukova dumka", Kiev, 1966.

12 May 1967

Institute of Cybernetics AS UkrSSR