BIPARTITE SET OF RANDOM EVENTS METHOD APPLICATION IN FACTORIAL ANALYSIS OF COMPLEX SYSTEMS

Irina Baranova

Siberian Federal University, Krasnoyarsk, Russia, *irinabar@yandex.ru*

In the paper is offered the new method of factorial analysis [1] based on the bipartite set of random events method [2]. It allows to survey complex system describing by numeric and set data on eventological ─ events level. It is considered an application of suggested method for the solution of practical problem ─ problem of determining indices whose have mostly influence on the public health.

There are many methods of classical factorial analysis dedicated to finding the connection between the random variables. New eventological method allows to working with polytypic data describing system's behavior and comparison of system's elements.

In some fields of science and practical activities different researches result to solving the problems of system analysis. In paper it is considered a situation, when the one part of the events describing the complex system's behaviour is numerical and the second part is sets. Eventology suggests analyzing the random events, whose describe of system's behaviour. The basic difficulty in analysis of such complex systems lies in the fact that number of all possible events is big and the data describing system's behavior is polytipic. This problem especially actual for applied fields of science, whose are bound up with analysis of social, economic and natural systems. There are medicine, ecology, biology, actuary, finances, insurance, sociology and others.

In work [2] was suggested the bipartite set of random events method, in which each system's element represents as bipartite set of random events. The first part of one corresponds the random variables, and second part - sets. The basic idea of this method concludes in reduction a analysis of system's elements to analysis of corresponding them bipartite sets of events. It was offered to use the probability of Minkovsky symmetry difference set-operation as distance between bipartite sets of random events.

The new approach for factorial analysis based on foregoing method concludes in next actions:

1. find difference between system's elements on each parameter (separately from first and second part of set);
2. arrange parameters in order of decreasing probabilities;
3. select parameters having probabilities more then necessary cutting off level.

In work also is considered an application of suggested method for the solution of practical problem ─ problem of determining indices whose have mostly influence on the public health.

Bipartite set of random elements

In paper we survey complex system, whose behavior is describing by numeric and set data. Then the results of observation for the behavior of researching object is a set, which consists of random variables and random sets. As it is known, each random variable and each random finite set, follow Kolmogorov, can be considered as random element's realization [2].

Let (Ω, F, P) is probabilistic space, where *F* is set of events, $x, y \in F$ ─ random events, ***P*** ─ probability measure. The measurable reflection

$$K : (\Omega, F, P) \to \left(2^{X}, 2^{2^{X}}\right),$$

which possible values are subsets of the finite set of events F, is called a random set of events *K*, defined under the given finite set of events $X \in F$ by probability distribution

$$p(X) = P(K = X),\ X \in 2^{X}.$$

As stated above, our general situation, when the one part of the results of observation for behavior of researching object are numerical and the second part — sets, can be described as the set of the random elements.

The set of the random elements $\{\xi, \mathrm{K}\}$ we will call the *bipartite set of random elements*. It can be defined in that way:

$$\{\xi, \mathrm{K}\} = \xi \cup K = \{\xi_a, a \in A, K_\beta, \beta \in B\}.$$

This set of random elements separates on the two parts. The first part is the random variables ξ, the second part is the random sets *K*. Let *A* is the indices set of a random variables, *B* — indices set of a random sets.

Bipartite set of random events

Now we considered the bipartite set of random events corresponded the bipartite set of random elements, defines above.

Let $\{\xi_a, a \in A\}$ are random variables with finite set of possible values

$$\Re_a = \{r_{a_1}, \ldots, r_{a_{Na}}\} \subset \mathrm{R},\ \ a \in \mathrm{A}.$$

To each random variables can be put in correspondence set of events

$$\xi_a \Rightarrow Y_a = \{Y_a(r_a), r_a \in \Re_a\}.$$

The event $Y_a(r_a) = \{\omega : \xi_a(\omega) \le r_a\}$ is the event from definition of distribution function of random variable and a set with inserting structure of dependences.

To first part — random variables can be put in correspondence set of events:

$$\xi \Rightarrow Y = \bigcup_{a \in A} Y_a.$$

To each random set of events K_β can be put in correspondence the finite set of events X_β:

$$K_\beta \Leftrightarrow X_\beta.$$

To second part — random sets of events *K* can be put in correspondence common set of events X:

$$K \Leftrightarrow \mathrm{X} = \sum_{\beta \in B} X_\beta,\ \beta \in B.$$

Bipartite set of random events can be defined in that way:

$$\{Y, \mathrm{X}\} = \{Y_a, X_\beta, a \in A, \beta \in B\}.$$

It is suggested for the bipartite set of random elements $\{\xi, \mathrm{K}\}$ put in correspondence the bipartite set of random events $\{Y, \mathrm{X}\}$.

The bipartite set of random events method

The basic idea of bipartite set of random events method [2] concludes in reduction an analysis of system's elements to analysis of corresponding bipartite sets of events.

Let system consists of *n* elements. Each element describes by parameters, one part of them is numerical and the second part is sets.

Let $s^+ = \{Y^+, \mathrm{X}^+\}, s^1 = \{Y^1, \mathrm{X}^1\}, \ldots, s^n = \{Y^n, \mathrm{X}^n\}$ — bipartite sets of random events describing the state of the "best" ideal system's element (the best values for all parameters) and *n* system's element.

The comparison between system's elements is difficult, but it is offered the reduction a system's elements to corresponding bipartite sets of events. If we known the eventological distributions of the bipartite sets of events then we can compare them by using the Minkovsky set-operations [2].

The Minkovsky set-operations

The operation under events from the each part consisted the bipartite sets of events is called an arbitrary Minkovsky set-operation. For example, the Minkovsky symmetry difference set-operation of two bipartite sets of random events s^1 and s^2:

$$s^1(\Delta)s^2 = \left\{Y_a^1(\Delta)Y_a^2, X_\beta^1(\Delta)X_\beta^2,\ X_\chi \subseteq \mathrm{X}_\beta, a \in A, \beta \in B\right\} =$$
$$= \left\{Y_a^1(r_a)(\Delta)Y_a^2(r_a), X_\beta^1(\Delta)X_\beta^2,\ r_a \in \Re_a,\ X_\chi \subseteq \mathrm{X}_\beta, a \in A, \beta \in B\right\}.$$

The probability of the Minkovsky symmetry difference set-operation:

$$\mathrm{P}\left(s^1(\Delta)s^2\right) = \frac{1}{|A|}\sum_{a\in A}\frac{1}{|Y_a|}\sum_{r_a\in\Re_a}\mathrm{P}\left(Y_a^1(r_a)\Delta Y_a^2(r_a)\right) +$$
$$+\frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|X_\beta|}\sum_{X_\beta\subseteq \mathrm{X}_\beta}\mathrm{P}\left(X_\beta^1 \Delta X_\beta^2\right).$$

We proved that probability of the Minkovsky symmetry difference set-operation of two bipartite sets of random events is the pseudometrics between their [2]. And because of that it is suggested to use the probability of the Minkovsky symmetry difference set-operation as a distance between bipartite sets of random events.

The bipartite set method application in factorial analysis

The comparison between two system's elements s^i and s^j carry out by finding the probability of the Minkovsky symmetry difference set-operation:

$$\mathrm{P}\left(s^i(\Delta)s^j\right) = \frac{1}{|A|}\sum_{a\in A}\frac{1}{|Y_a|}\sum_{r_a\in\Re_a}\mathrm{P}\left(Y_a^i(r_a)\Delta Y_a^j(r_a)\right) +$$
$$+\frac{1}{|B|}\sum_{\beta\in B}\frac{1}{|X_\beta|}\sum_{X_\beta\subseteq \mathrm{X}_\beta}\mathrm{P}\left(X_\beta^i \Delta X_\beta^j\right).$$

Let us consider the parameters mostly influence on the difference between system's elements, if they have the largest meanings of Minkovsky symmetry difference set-operation between elements. Thus, new approach for factorial analysis consists of these actions:

1. find difference between system's elements on each parameter (separately from first and second part of set);
2. arrange parameters in order of decreasing probabilities;
3. select parameters having probabilities more then necessary cutting off level.

The problem of determining indices whose mostly influence on the population health of the Krasnoyarsk land

It is considered the system of population health districts of Krasnoyarsk land. The districts are the elements of this system. In statistics of population health are 50 districts (42 administrative districts and 8 cities). The population health of each districts describes by 74 parameters. The 72 health parameters are numerical. There are climatic (average monthly air temperature, average monthly wind speed, climate rigidity, etc), economic (average monthly income per head, unemployment level, average living wage, etc) and anthropogeneous parameters (gross wasting of polluting substances). Two last parameters are the sets of events: the set of chemical elements polluting district's ground and the set of diseases, whose has a population of district (their values was described above in section of bipartite set of random elements). It is considered an application of suggested method for the solution of practical problem ─ problem of determining indices whose have mostly influence on the public health.

Eventological distribution of each district estimates from the statistics. Then it is finding the probabilities of the Minkovsky symmetry difference set-operation by the foregoing formulae. After that we arranged parameters in order of decreasing probabilities and select those indices having probabilities more then 0.2.

Scientific novelty of this paper concludes in next theses:

1. For first time factorial analysis of complex system carried out on eventological level, on which a analysis of system's elements reduce to analysis of corresponding them bipartite sets of events.
2. Suggested method allows to simplify working with polytypic data describing system's behavior and comparison of system's elements.
3. Suggested method finds number of factors describing correlation connections between parameters and meaning of parameters influence on the difference between elements.

Literature

1. I.V. Baranova. The application of bipartite set method application in factorial analysis of complex system (In Russian) / Bulletin of Krasnoyarsk State University, Krasnoyarsk (2006), No. 1, pp. 142-152.
2. O.Yu. Vorob'ov, I.V. Baranova. The bipartite set method application in eventological analysis of complex system (In Russian) Siberian Federal University, Krasnoyarsk (2007), 132 p.